

The tragic ripples of  
an epic fraud *p. 636*

Insect pest profits from  
maize defenses *pp. 642 & 694*

Photoredox activation  
of methane *pp. 647 & 668*

# Science

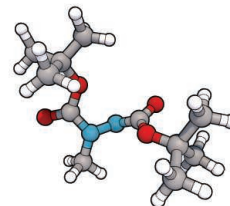
\$15  
17 AUGUST 2018  
[sciencemag.org](http://sciencemag.org)

AAAS



## ROAD MAP FOR WHEAT

Ordered sequence will  
speed research *pp. 635, 661, & 662*

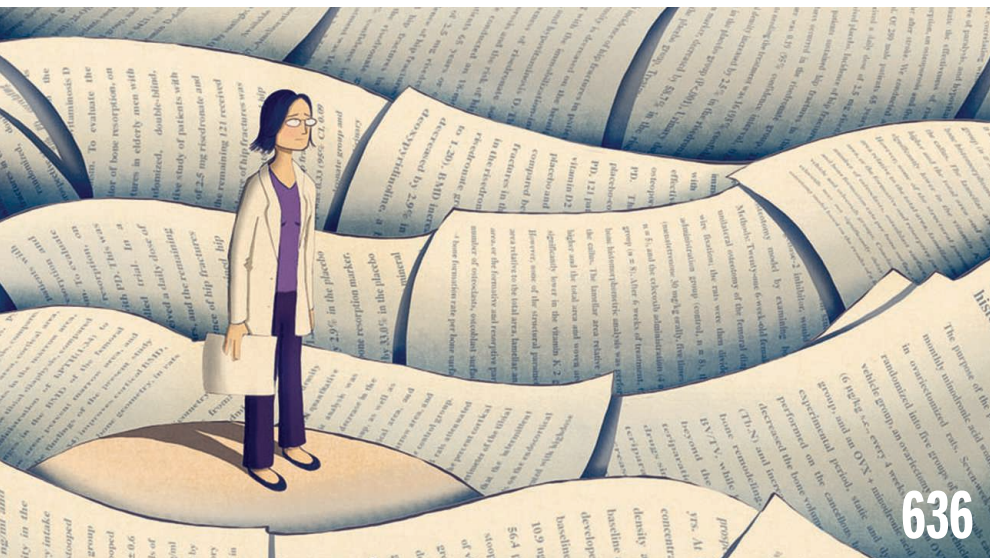


647 & 668

Slicing simple hydrocarbons

# CONTENTS

17 AUGUST 2018 • VOLUME 361 • ISSUE 6403



## NEWS

### IN BRIEF

**628** News at a glance

### IN DEPTH

#### **630** NEW CASE OF ALLEGED BULLYING ROCKS THE MAX PLANCK SOCIETY

Colleagues say empathy researcher Tania Singer created an atmosphere of fear at her lab *By K. Kupferschmidt*

#### **631** CRITICS PAN EPA PLAN FOR WEIGHING TOXIC CHEMICAL RISKS

Guidance could allow regulators to exclude key studies, researchers argue *By V. Zainzinger*

#### **632** AI TAKES ON VIDEO GAMES IN QUEST FOR COMMON SENSE

Bots seek to hone strategy skills in matchup with top *Dota 2* players *By M. Hutson*

#### **634** STUDY REIGNITES DEBATE ABOUT WHEN THERA BLEW ITS TOP

Radiocarbon curve suggests archaeological data were right *By L. Wade*

SCIENCE ADVANCES RESEARCH ARTICLE BY C. L. PEARSON ET AL. 10.1126/SCIADV.AAR8241; PODCAST

#### **635** DETAILED GENOME MAPS PATHS TO BETTER WHEAT

Chromosome-by-chromosome sequence helps navigate crop's genomic complexity *By E. Pennisi*

► RESEARCH ARTICLES P. 661 & 662; SCIENCE ADVANCES RESEARCH ARTICLE BY A. JUHÁSZ ET AL. 10.1126/SCIADV.AAR8602

### FEATURE

#### **636** TIDE OF LIES

The researcher at the center of an epic scientific fraud remains an enigma to the scientists who exposed him

*By K. Kupferschmidt*

## INSIGHTS

### PERSPECTIVES

#### **642** PLANT NUTRIENT ACQUISITION ENTICES HERBIVORE

Maize defense metabolites enable iron uptake and attract herbivore attention *By D. J. Kliebenstein*

► REPORT P. 694

#### **643** CHOOSING THE RIGHT INPUT IN CELL SIGNALING

A decoding module discriminates ligands and signaling activation in the Wnt pathway *By K. Kim and L. Goentoro*

► RESEARCH ARTICLE P. 663

#### **645** CAGED CLUSTERS SHINE BRIGHTER

The structure of small silver clusters confined in zeolites helps to explain their luminescent properties

*By M. Quintanilla and L. M. Liz-Marzán*

► REPORT P. 686

#### **646** PHASE SEPARATION FOCUSES DNA SENSING

Liquid droplets form upon recognition of DNA to stimulate innate immune responses *By A. Ablasser*

► REPORT P. 704

#### **647** PHOTOCATALYTIC UPGRADING OF NATURAL GAS

Cerium salts catalyze formation of value-added molecules from methane and small alkanes *By M. Kanai*

► REPORT P. 668

#### **648** TURNING UP TRANSLATION IN FRAGILE X SYNDROME

Fragile X mental retardation protein promotes translation, contrary to its assumed function *By S. Aryal and E. Klann*

► REPORT P. 709

### POLICY FORUM

#### **650** GLOBAL WARMING POLICY: IS POPULATION LEFT OUT IN THE COLD?

Population policies offer options to lessen climate risks

*By J. Bongaarts and B. C. O'Neill*

### BOOKS ET AL.

#### **653** AN ISLAND ON THE BRINK

Convinced erosion, not climate change, threatens their home, a community grapples with an uncertain future

*By A. J. Miller*

#### **654** OUR LOOMING LEAD PROBLEM

Neglect, poor planning, and bad decisions led to Flint's water crisis. It could easily happen again.

*By F. R. Davis*

### LETTERS

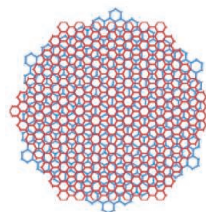
#### **655** HARASSMENT CHARGES: ENOUGH HIMPATRY

*By J. Zelikova et al.*

#### **655** HARASSMENT CHARGES: JOURNALISTS' ROLE

*By J. Duffy*





690

Twisting a  
graphene sandwich

# CONTENTS

17 AUGUST 2018 • VOLUME 361 • ISSUE 6403

## 655 HARASSMENT CHARGES: INJUSTICE DONE?

By A. Moya et al.

## 656 HARASSMENT CHARGES: METOO BUT DUE PROCESS

By K. R. Monroe

## RESEARCH

### IN BRIEF

657 From *Science* and other journals

### REVIEW

#### 660 OPTICAL IMAGING

Quantum-inspired computational  
imaging Y. Altmann et al.

REVIEW SUMMARY; FOR FULL TEXT:

dx.doi.org/10.1126/science.aat2298

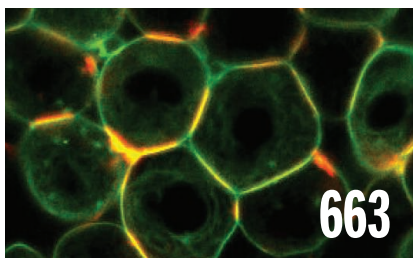
### RESEARCH ARTICLES

#### WHEAT GENOME

661 Shifting the limits in wheat  
research and breeding using a  
fully annotated reference genome  
*International Wheat Genome  
Sequencing Consortium (IWGSC)*  
RESEARCH ARTICLE SUMMARY; FOR FULL  
TEXT: dx.doi.org/10.1126/science.aar7191

662 The transcriptional landscape of  
polyploid wheat  
R. H. Ramírez-González et al.  
RESEARCH ARTICLE SUMMARY; FOR FULL  
TEXT: dx.doi.org/10.1126/science.aar6089

► NEWS STORY P. 635; *SCIENCE ADVANCES*  
RESEARCH ARTICLE BY A. JUHÁSZ ET AL.  
10.1126/SCIADV.AAR8602



#### 663 DEVELOPMENTAL SIGNALING

A molecular mechanism for Wnt ligand-  
specific signaling M. Eubelen et al.

RESEARCH ARTICLE SUMMARY; FOR FULL TEXT:

dx.doi.org/10.1126/science.aat1178

► PERSPECTIVE P. 643

### REPORTS

#### 664 ORGANIC CHEMISTRY

Stereodivergent synthesis of  
1,4-dicarbonyls by traceless  
charge-accelerated sulfonium  
rearrangement D. Kaldre et al.

#### 668 ORGANIC CHEMISTRY

Selective functionalization of methane,  
ethane, and higher alkanes by cerium  
photocatalysis A. Hu et al.

► PERSPECTIVE P. 647

#### 672 ACTIVE MATTER

Collective clog control: Optimizing  
traffic flow in confined biological and  
robophysical excavation J. Aguilar et al.

#### 677 HIGH-PRESSURE PHYSICS

Insulator-metal transition in  
dense fluid deuterium  
P. M. Celliers et al.

#### 682 MEMBRANES

3D printed polyamide membranes  
for desalination M. R. Chowdhury et al.

#### 686 PHOTOLUMINESCENCE

Origin of the bright photo-  
luminescence of few-atom silver  
clusters confined in LTA zeolites  
D. Grandjean et al.

► PERSPECTIVE P. 645

#### 690 SOLID-STATE PHYSICS

Twistable electronics with dynamically  
rotatable heterostructures  
R. Ribeiro-Palau et al.

#### 694 PLANT SCIENCE

Plant iron acquisition strategy  
exploited by an insect herbivore  
L. Hu et al.

► PERSPECTIVE P. 642

#### 697 CLIMATE

Abrupt cloud clearing of marine  
stratocumulus in the subtropical  
southeast Atlantic S. E. Yuter et al.

#### 701 RNA METABOLISM

Mixed tailing by TENT4A and TENT4B  
shields mRNA from rapid deadenylation  
J. Lim et al.

#### 704 IMMUNOLOGY

DNA-induced liquid phase condensation  
of cGAS activates innate immune  
signaling M. Du et al.

► PERSPECTIVE P. 646

#### 709 TRANSLATIONAL CONTROL

Fragile X mental retardation 1 gene  
enhances the translation of large  
autism-related proteins  
E. J. Greenblatt and A. C. Spradling

► PERSPECTIVE P. 648

### DEPARTMENTS

#### 627 EDITORIAL

Don't lose race-conscious policies  
By Liliana M. Garces

#### 718 WORKING LIFE

The universe in a classroom  
By Jeffrey Maloy

### ON THE COVER



Spikes of bread wheat  
(*Triticum aestivum*),  
a hybrid of three  
distinct genomes.  
Scientists have recently  
completed sequencing  
of a high-quality refer-  
ence wheat genome  
and identified how the

three ancestral genomes work at a genetic  
level to form a staple of human nutrition.  
See pages 635, 661, and 662.

Photo: Christian Vidal/Science Source

Science Staff ..... 626  
New Products ..... 713  
Science Careers ..... 714

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2018 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): \$165 (\$74 allocated to subscription). Domestic institutional subscription (51 issues): \$1808; Foreign postage extra: Mexico, Caribbean (surface mail) \$55; other countries (air assist delivery): \$89. First class, airmail, student, and emeritus rates with GST available upon request. Canadian rates with GST available upon request. GST #R125488122. Publications Mail Agreement Number 1069624. Printed in the U.S.A. Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. Postmaster: Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090-6178. Single-copy sales: \$15 each plus shipping and handling; bulk rate on request. Authorization to reproduce material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and others who use Copyright Clearance Center (CCC) Pay-Per-Use services provided that \$35.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for Science is 0036-8075. Science is indexed in the Reader's Guide to Periodical Literature and in several specialized indexes.

**Editor-in-Chief** Jeremy Berg

**Executive Editor** Monica M. Bradford **News Editor** Tim Appenzeller

**Deputy Editors** Lisa D. Chong, Andrew M. Sugden(UK), Valda J. Vinson, Jake S. Yeston

## Research and Insights

**DEPUTY EDITOR, EMERITUS** Barbara R. Jasny **SR. EDITORS** Gemma Alderton(UK), Caroline Ash(UK), Julia Fahrenkamp-Uppenbrink(UK), Pamela J. Hines, Stella M. Hurlley(UK), Paula A. Kiberstis, Marc S. Lavine(Canada), Steve Mao, Ian S. Osborne(UK), Beverly A. Purnell, L. Bryan Ray, H. Jesse Smith, Jelena Stajic, Peter Stern(UK), Phillip D. Szuromi, Sacha Vignieri, Brad Wible, Laura M. Zahn **ASSOCIATE EDITORS** Michael A. Funk, Brent Grocholski, Priscilla N. Kelly, Tage S. Rai, Seth Thomas Scanlon(UK), Keith T. Smith(UK) **ASSOCIATE BOOK REVIEW EDITOR** Valerie B. Thompson **LETTERS EDITOR** Jennifer Sills **LEAD CONTENT PRODUCTION EDITORS** Harry Jach, Lauren Kmec **CONTENT PRODUCTION EDITORS** Amelia Beyna, Jeffrey E. Cook, Amber Esplin, Chris Filiatreau, Cynthia Howe, Catherine Wolner **SR. EDITORIAL COORDINATORS** Carolyn Kyle, Beverly Shields **EDITORIAL COORDINATORS** Aneera Dobbins, Joi S. Granger, Jeffrey Hearn, Lisa Johnson, Maryrose Madrid, Jerry Richardson, Alice Whaley(UK), Anita Wynn **PUBLICATIONS ASSISTANTS** Ope Martins, Nida Masiulis, Dona Mathieu, Ronmel Navas, Hilary Stewart(UK), Alana Warnke, Brian White **EXECUTIVE ASSISTANT** Jessica Slater **ADMINISTRATIVE SUPPORT** Janet Clements(UK), Jessica Waldo(UK), Ming Yang (UK)

## News

**NEWS MANAGING EDITOR** John Travis **INTERNATIONAL EDITOR** Martin Enserink **DEPUTY NEWS EDITORS** Elizabeth Culotta, David Grimm, Eric Hand, David Malakoff, Leslie Roberts **SR. CORRESPONDENTS** Daniel Clery(UK), Jeffrey Mervis, Elizabeth Pennisi **ASSOCIATE EDITORS** Jeffrey Brainerd, Catherine Maticic **NEWS WRITERS** Adrian Cho, Jon Cohen, Jennifer Couzin-Frankel, Jocelyn Kaiser, Kelly Servick, Robert F. Service, Erik Stokstad(Cambridge, UK), Paul Voosen, Meredith Wadman **INTERNS** Katie Langin, Frankie Schembri, Matt Warren **CONTRIBUTING CORRESPONDENTS** Warren Cornwall, Ann Gibbons, Mara Hvistendahl, Sam Kean, Eli Kintisch, Kai Kupferschmidt(Berlin), Andrew Lawler, Matt Leslie, Eliot Marshall, Virginia Morell, Dennis Normile(Shanghai), Charles Pillar, Tania Rabesandratana(London), Emily Underwood, Gretchen Vogel(Berlin), Lizzie Wade(Mexico City) **CAREERS** Donisha Adams, Rachel Bernstein(Editor) **COPY EDITORS** Julia Cole (Senior Copy Editor), Cyra Master (Copy Chief) **ADMINISTRATIVE SUPPORT** Meagan Weiland

**Executive Publisher** Rush D. Holt

**Publisher** Bill Moran **Chief Digital Media Officer** Josh Freeman

**DIRECTOR, BUSINESS STRATEGY AND PORTFOLIO MANAGEMENT** Sarah Whalen **DIRECTOR, PRODUCT AND CUSTOM PUBLISHING** Will Schweitzer **MANAGER, PRODUCT DEVELOPMENT** Hannah Heckner **BUSINESS SYSTEMS AND FINANCIAL ANALYSIS** Director Randy Yi **DIRECTOR, BUSINESS OPERATIONS & ANALYST** Eric Knott **ASSOCIATE DIRECTOR, PRODUCT MANAGEMENT** Kris Bishop **ASSOCIATE DIRECTOR, INSTITUTIONAL LICENSING** SALE Geoffrey Worton **SENIOR SYSTEMS ANALYST** Nicole Mehmedovich **SENIOR BUSINESS ANALYST** Cory Lipman **MANAGER, BUSINESS OPERATIONS** Jessica Tierney **BUSINESS ANALYSTS** Meron Kebede, Sandy Kim, Jourdan Stewart **FINANCIAL ANALYST** Julian Iriarte **ADVERTISING SYSTEM ADMINISTRATOR** Tina Burks **SALES COORDINATOR** Shirley Young **DIRECTOR, COPYRIGHT, LICENSING, SPECIAL PROJECTS** Emilie David **DIGITAL PRODUCT ASSOCIATE** Michael Hardesty **RIGHTS AND PERMISSIONS ASSOCIATE** Elizabeth Sandler **RIGHTS, CONTRACTS, AND LICENSING ASSOCIATE** Lili Catlett **RIGHTS & PERMISSIONS ASSISTANT** Alexander Lee

**MARKETING MANAGER, PUBLISHING** Shawana Arnold **SENIOR ART ASSOCIATES** Paula Fry **ART ASSOCIATE** Kim Huynh

**DIRECTOR, INSTITUTIONAL LICENSING** Iqo Edim **ASSOCIATE DIRECTOR, RESEARCH & DEVELOPMENT** Elisabeth Leonard **SENIOR INSTITUTIONAL LICENSING MANAGER** Ryan Rexroth **INSTITUTIONAL LICENSING MANAGERS** Marco Castellani, Chris Murawski **SENIOR OPERATIONS ANALYST** Lana Guz **MANAGER, AGENT RELATIONS & CUSTOMER SUCCESS** Judy Lillibridge

**WEB TECHNOLOGIES TECHNICAL DIRECTOR** David Levy **PORTFOLIO MANAGER** Trista Smith **PROJECT MANAGER** Dean Robbins **DEVELOPERS** Liana Birke, Elissa Heller, Ryan Jensen

**DIGITAL MEDIA DIRECTOR OF ANALYTICS** Enrique Gonzales **MULTIMEDIA MANAGER** Sarah Crespi **MANAGING WEB PRODUCER** Kara Estelle-Powers **DIGITAL PRODUCER** Jessica Hubbard **VIDEO PRODUCER** Chris Burns **SOCIAL MEDIA PRODUCER** Brice Russ

**DIGITAL/PRINT STRATEGY MANAGER** Jason Hillman **QUALITY TECHNICAL MANAGER** Marcus Spiegler **DIGITAL PRODUCTION MANAGER** Lisa Stanford **ASSISTANT MANAGER DIGITAL/PRINT** Rebecca Doshi **SENIOR CONTENT SPECIALISTS** Steve Forrester, Antoinette Hodal, Lori Murphy, Anthony Rosen **CONTENT SPECIALISTS** Jacob Hedrick, Kimberley Oster

**DESIGN DIRECTOR** Beth Rakouskas **DESIGN MANAGING EDITOR** Marcy Atarod **SENIOR DESIGNER** Chrystal Smith **DESIGNER** Christina Aycock **GRAPHICS MANAGING EDITOR** Alberto Cuadra **GRAPHICS EDITOR** Nirja Desai **SENIOR SCIENTIFIC ILLUSTRATORS** Valerie Altounian, Chris Bickel, Katharine Sutfill **SCIENTIFIC ILLUSTRATOR** Alice Kitterman **INTERACTIVE GRAPHICS EDITOR** Jia You **SENIOR GRAPHICS SPECIALISTS** Holly Bishop, Nathalie Cary **PHOTOGRAPHY MANAGING EDITOR** William Douthitt **PHOTO EDITOR** Emily Petersen **IMAGE RIGHTS AND FINANCIAL MANAGER** Jessica Adams

**SENIOR EDITOR, CUSTOM PUBLISHING** Sean Sanders: 202-326-6430 **ASSISTANT EDITOR, CUSTOM PUBLISHING** Jackie Oberst: 202-326-6463 **ASSOCIATE DIRECTOR, BUSINESS DEVELOPMENT** Justin Sawyers: 202-326-7061 **science\_advertising@aaas.org** **ADVERTISING PRODUCTION OPERATIONS MANAGER** Deborah Tompkins **SR. PRODUCTION SPECIALIST/GRAPHIC DESIGNER** Amy Hardcastle **SR. TRAFFIC ASSOCIATE** Christine Hall **DIRECTOR OF BUSINESS DEVELOPMENT AND ACADEMIC PUBLISHING RELATIONS, ASIA** Xiaoying Chu: +86-131 6136 3212, xchu@aaas.org **COLLABORATION/CUSTOM PUBLICATIONS/JAPAN** Adarsh Sandhu + 81532-81-5142 asandhu@aaas.org **EAST COAST/E. CANADA** Laurie Faraday: 508-747-9395, FAX 617-507-8189 **WEST COAST/W. CANADA** Lynne Stickrod: 415-931-9782, FAX 415-520-6940 **MIDWEST** Jeffrey Dembski: 847-498-4520 x3005, Steven Loerch: 847-498-4520 x3006 **UK EUROPE/ASIA** Roger Goncalves: TEL/FAX +41 43 243 1358 **JAPAN** Kaoru Sasaki (Tokyo): +81 (3) 6459 4174 ksasaki@aaas.org

**GLOBAL SALES DIRECTOR ADVERTISING AND CUSTOM PUBLISHING** Tracy Holmes: +44 (0) 1223 326525 **CLASSIFIED** advertise@sciencecareers.org **SALES MANAGER, US, CANADA AND LATIN AMERICA** SCIENCE CAREERS Claudia Paulsen-Young: 202-326-6577 **EUROPE/ROW SALES** Sarah Lelarge **SALES ADMIN ASSISTANT** Kelly Grace +44 (0)1223 326528 **JAPAN** Miyuki Tani(Osaka): +81 (6) 6202 6272 mtani@aaas.org **CHINA/TAIWAN** Xiaoying Chu: +86-131 6136 3212, xchu@aaas.org **GLOBAL MARKETING MANAGER** Allison Pritchard **DIGITAL MARKETING ASSOCIATE** Aimee Aponte

**AAAS BOARD OF DIRECTORS, CHAIR** Susan Hockfield **PRESIDENT** Margaret A. Hamburg **PRESIDENT-ELECT** Steven Chu **TREASURER** Carolyn N. Ainslie **CHIEF EXECUTIVE OFFICER** Rush D. Holt **BOARD** Cynthia M. Beall, May R. Berenbaum, Rosina M. Bierbaum, Kaye Husbands Fealing, Stephen P.A. Fodor, S. James Gates, Jr., Michael S. Gazzaniga, Laura H. Greene, Robert B. Millard, Mercedes Pascual, William D. Provine

**SUBSCRIPTION SERVICES** For change of address, missing issues, new orders and renewals, and payment questions: 866-434-AAAS (2227) or 202-326-6417, FAX 202-842-1065. Mailing addresses: AAAS, P.O. Box 96178, Washington, DC 20090-6178 or AAAS Member Services, 1200 New York Avenue, NW, Washington, DC 20005

**INSTITUTIONAL SITE LICENSES** 202-326-6730 **REPRINTS:** Author Inquiries 800-635-7181 **COMMERCIAL INQUIRIES** 803-359-4578 **PERMISSIONS** 202-326-6765, permissions@aaas.org **AAAS Member Central Support** 866-434-2227 [www.aaas.org/membercentral](http://www.aaas.org/membercentral)

Science serves as a forum for discussion of important issues related to the advancement of science by publishing material on which a consensus has been reached as well as including the presentation of minority or conflicting points of view. Accordingly, all articles published in Science—including editorials, news and comment, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.

**INFORMATION FOR AUTHORS** See [www.sciencemag.org/authors/science-information-authors](http://www.sciencemag.org/authors/science-information-authors)

## BOARD OF REVIEWING EDITORS (Statistics board members indicated with \$)

Adriano Aguzzi, *U. Hospital Zürich*  
Takuzo Aida, *U. of Tokyo*  
Leslie Aiello, *Wenner-Gren Foundation*  
Judith Allen, *U. of Manchester*  
Sebastian Amigorena, *Institut Curie*  
Meinrat O. Andrae, *Max Planck Inst. Mainz*  
Paola Ariotti, *Harvard U.*  
Johan Auwerx, *EPFL*  
David Awschalom, *U. of Chicago*  
Clare Baker, *U. of Cambridge*  
Nenad Ban, *ETH Zürich*  
Franz Bauer, *Pontificia Universidad Católica de Chile*  
Ray H. Baughman, *U. of Texas at Dallas*  
Carlo Beenakker, *Leiden U.*  
Kamran Behnia, *ESPCI*  
Yasmine Belkaid, *NIAD, NIH*  
Philip Benfey, *Duke U.*  
Gabriele Bergers, *ViB*  
Bradley Bernstein, *Massachusetts General Hospital*  
Peer Bork, *EMBL*  
Chris Bowler, *École Normale Supérieure*  
Ian Boyd, *U. of St. Andrews*  
Emily Brodsky, *U. of California, Santa Cruz*  
Ron Brookmeyer, *U. of California, Los Angeles (\$)*  
Christian Büchel, *UKE Hamburg*  
Dennis Burton, *The Scripps Res. Inst.*  
Carter Tribley Butts, *U. of California, Irvine*  
Gyorgy Buzsaki, *New York U. School of Medicine*  
Blanche Capel, *Duke U.*  
Mats Carlsson, *U. of Oslo*  
Ib Chorkendorff, *Denmark TU*  
James J. Collins, *MIT*  
Robert Cook-Deegan, *Arizona State U.*  
Lisa Coussens, *Oregon Health & Science U.*  
Alan Cowman, *Walter & Eliza Hall Inst.*  
Roberta Croce, *VU Amsterdam*  
Jeff L. Dangl, *U. of North Carolina*  
Tom Daniel, *U. of Washington*  
Chiara Daraio, *Caltech*  
Nicolas Dauphas, *U. of Chicago*  
Frans de Waal, *Emory U.*  
Stanislas Dehaene, *Collège de France*  
Robert Desimone, *MIT*  
Claude Desplan, *New York U.*  
Sandra Diaz, *Universidad Nacional de Córdoba*  
Dennis Discher, *U. of Penn.*  
Gerald W. Dorn II, *Washington U. in St. Louis*  
Jennifer A. Doudna, *U. of California, Berkeley*  
Bruce Dunn, *U. of California, Los Angeles*  
William Dunphy, *Caltech*  
Christopher Dye, *U. of Oxford*  
Todd Ehlers, *U. of Tübingen*  
Jennifer Elisseeff, *Johns Hopkins U.*  
Tim Elston, *U. of North Carolina at Chapel Hill*  
Barry Everitt, *U. of Cambridge*  
Vanessa Ezenwa, *U. of Georgia*  
Ernst Fehr, *U. of Zürich*  
Michael Feuer, *The George Washington U.*  
Toren Finkel, *U. of Pittsburgh Medical Ctr.*  
Kate Fitzgerald, *U. of Massachusetts*  
Peter Fratzl, *Max Planck Inst. Potsdam*  
Elaine Fuchs, *Rockefeller U.*  
Eileen Furlong, *EMBL*  
Jay Gallagher, *U. of Wisconsin*  
Daniel Geschwind, *U. of California, Los Angeles*  
Karl-Heinz Glassmeier, *TU Braunschweig*  
Ramon Gonzalez, *Rice U.*  
Elizabeth Grove, *U. of Chicago*  
Nicolas Gruber, *ETH Zürich*  
Kip Guy, *U. of Kentucky College of Pharmacy*  
Taekjip Ha, *Johns Hopkins U.*  
Christian Haass, *Ludwig Maximilians U.*  
Sharon Hammes-Schiffer, *U. of Illinois at Urbana-Champaign*  
Wolf-Dietrich Hardt, *ETH Zürich*  
Louise Harra, *U. College London*  
Michael Hasselmo, *Boston U.*  
Jian He, *Clemson U.*  
Martin Heimann, *Max Planck Inst. Jena*  
Carl-Philipp Heisenberg, *IST Austria*  
Ykä Helariutta, *U. of Cambridge*  
Janet G. Hering, *Eawag*  
Kai-Uwe Hinrichs, *U. of Bremen*  
David Hodell, *U. of Cambridge*  
Lora Hooper, *UT Southwestern Medical Ctr. at Dallas*  
Fred Hughson, *Princeton U.*  
Randall Hulet, *Rice U.*  
Auke Ijspeert, *EPFL*  
Akiko Iwasaki, *Yale U.*  
Stephen Jackson, *USGS and U. of Arizona*  
Seema Jayachandran, *Northwestern U.*  
Kai Johnsson, *EPFL*  
Peter Jonas, *Inst. of Science & Technology Austria*  
Matt Kaebberlein, *U. of Washington*  
William Kaelin Jr., *Dana-Farber Cancer Inst.*  
Daniel Kammen, *U. of California, Berkeley*  
Abby Kavner, *U. of California, Los Angeles*  
Masashi Kawasaki, *U. of Tokyo*  
V. Narry Kim, *Seoul Nat. U.*  
Robert Kingston, *Harvard Medical School*  
Etienne Kochlin, *École Normale Supérieure*  
Alexander Kolodkin, *Johns Hopkins U.*  
Thomas Langer, *U. of Cologne*  
Mitchell A. Lazar, *U. of Penn.*  
David Lazer, *Harvard U.*  
Stanley Lemon, *U. of North Carolina at Chapel Hill*  
Ottoline Leyser, *U. of Cambridge*  
Wendell Lim, *U. of California, San Francisco*  
Marcia C. Linn, *U. of California, Berkeley*  
Jianguo Liu, *Michigan State U.*  
Luis Liz-Marzán, *CIC biomaGUNE*  
Jonathan Losos, *Harvard U.*  
Ke Lu, *Chinese Acad. of Sciences*  
Christian Lüscher, *U. of Geneva*  
Fabienne Mackay, *U. of Melbourne*  
Anne Magurran, *U. of St. Andrews*  
Oscar Marín, *King's College London*  
Charles Marshall, *U. of California, Berkeley*  
Christopher Marx, *U. of Idaho*  
C. Robertson McClung, *Dartmouth College*  
Rodrigo Medellín, *U. of Mexico*  
Daham Medley, *London School of Hygiene & Tropical Med.*  
Jane Memmott, *U. of Bristol*  
Tom Misteli, *NCI, NIH*  
Yasushi Miyashita, *U. of Tokyo*  
Christian Morris, *U. of Edinburgh*  
Alison Motsinger-Reif, *NC State U. (\$)*  
Daniel Neumark, *U. of California, Berkeley*  
Kitty Nijmeijer, *TU Eindhoven*  
Helga Nowotny, *Austrian Council*  
Rachel O'Reilly, *U. of Warwick*  
Harry Orr, *U. of Minnesota*  
Pilar Ossorio, *U. of Wisconsin*  
Andrew Oswald, *U. of Warwick*  
Isabella Pagano, *Istituto Nazionale di Astrofisica*  
Margaret Palmer, *U. of Maryland*  
Steve Palumbi, *Stanford U.*  
Jane Parker, *Max Planck Inst. Cologne*  
Giovanni Parmigiani, *Dana-Farber Cancer Inst. (\$)*  
Samuel Pfaff, *Salk Inst. for Biological Studies*  
Matthieu Piel, *Institut Curie*  
Kathrin Plath, *U. of California, Los Angeles*  
Martin Plenio, *Ulm U.*  
Albert Polman, *FOM Institute for AMOLF*  
Elvira Poloczanska, *Alfred-Wegener-Inst.*  
Philippe Poulin, *CNRS*  
Jonathan Pritchard, *Stanford U.*  
David Randall, *Colorado State U.*  
Sarah Reisman, *Caltech*  
Félix A. Rey, *Institut Pasteur*  
Trevor Robbins, *U. of Cambridge*  
Amy Rosenzweig, *Northwestern U.*  
Mike Ryan, *U. of Texas at Austin*  
Mitinori Saitou, *Kyoto U.*  
Shimon Sakaguchi, *Osaka U.*  
Miquel Salmeron, *Lawrence Berkeley Nat. Lab*  
Nitin Samarth, *Penn. State U.*  
Jürgen Sandkühler, *Medical U. of Vienna*  
Alexander Schier, *Harvard U.*  
Wolfram Schlenker, *Columbia U.*  
Susannah Scott, *U. of California, Santa Barbara*  
Vladimir Shaleev, *Purdue U.*  
Beth Shapiro, *U. of California, Santa Cruz*  
Jay Shendure, *U. of Washington*  
Brian Shoichet, *U. of California, San Francisco*  
Robert Siliciano, *Johns Hopkins U. School of Medicine*  
Uri Simonsohn, *U. of Penn.*  
Lucia Sivilotti, *U. College London*  
Alison Smith, *John Innes Centre*  
Richard Smith, *U. of North Carolina at Chapel Hill (\$)*  
Mark Smyth, *QIMR Berghofer*  
Pam Soltis, *U. of Florida*  
John Speakman, *U. of Aberdeen*  
Tara Spire-Jones, *U. of Edinburgh*  
Allan C. Spradling, *Carnegie Institution for Science*  
Eric Steig, *U. of Washington*  
Paula Stephan, *Georgia State U.*  
V. S. Subrahmanian, *U. of Maryland*  
Ira Tabas, *Columbia U.*  
Sarah Teichmann, *U. of Cambridge*  
Shubha Tole, *Tata Inst. of Fundamental Research*  
Wim van der Putten, *Netherlands Inst. of Ecology*  
Bert Vogelstein, *Johns Hopkins U.*  
David Wallach, *Weizmann Inst. of Science*  
Jane-Ling Wang, *U. of California, Davis (\$)*  
David Waxman, *Fudan U.*  
Jonathan Weissman, *U. of California, San Francisco*  
Chris Wickle, *U. of Missouri (\$)*  
Terrie Williams, *U. of California, Santa Cruz*  
Ian A. Wilson, *The Scripps Res. Inst. (\$)*  
Timothy D. Wilson, *U. of Virginia*  
Yu Xie, *Princeton U.*  
Jan Zaenen, *Leiden U.*  
Kenneth Zaret, *U. of Penn. School of Medicine*  
Jonathan Zehr, *U. of California, Santa Cruz*  
Maria Zuber, *MIT*



# Don't lose race-conscious policies

Using race as a factor in admission decisions by U.S. colleges and universities—so-called affirmative action—has once again become a lightning rod for debate. Last month, several universities defended Harvard University against a lawsuit that attacks its use of race in student admissions. The backdrop of this case, and a similar one against the University of North Carolina–Chapel Hill, is even more disturbing. Also last month, the U.S. Department of Justice (DOJ) and Department of Education (DOE), both under leadership appointed by the Trump administration, encouraged “race-neutral” admissions practices, rolling back previous guidance for achieving diversity by these very agencies. The U.S. Supreme Court has repeatedly allowed higher education institutions to consider race as one factor, among many, in admissions decisions. The rollback does not change that precedent.

Two summers ago, the Supreme Court ruled that colleges and universities can use race in admissions decisions. This *Fisher v. University of Texas* decision added to a 40-year-old line of Supreme Court precedent determining that the policy helped postsecondary institutions attain the educational benefits of diversity. In 2011, 2013, and 2016, the DOJ and DOE issued guidance documents clarifying the implications of this precedent for higher education practitioners and administrators. Although not legally binding, these documents outlined the legal framework and actions that institutions could take to achieve diversity and advance their educational mission.

The 2018 decision to rescind this guidance follows action last year by the DOJ to redirect its resources toward investigating claims of discrimination at institutions that employ race-conscious admissions. These actions seek to deny the constitutionally endorsed tool that universities use to diversify and drive innovation—actions that improve public well-being and scientific inquiry. Is it a coincidence that the DOJ and DOE rollback was announced at a time when legal pro-

ceedings on two affirmative action cases are taking place?

As a legal scholar and researcher focused on access and equity in higher education, I have represented hundreds of social science researchers who have defended affirmative action before the Supreme Court. Those briefs joined over 100 other friend-of-the-court briefs filed in support of affirmative action in the *Fisher* case. In the face of a potential threat of legal action by the DOJ, it would be understandable for institutions to consider abandoning race-conscious policies. But doing so will have a dramatic and devastating effect on student body diversity and the

racial climate of college campuses more generally. As public postsecondary institutions in states that ban affirmative action (so far, in eight states) have already experienced, the inability to consider race in admissions leads to substantial drops in the diversity of student bodies. These declines have taken place at public selective undergraduate institutions; in graduate fields of engineering, natural, and social sciences; and in law and medical schools. At public medical schools across six states with bans, for example, the share of enrolled students who are students of color dropped from 18.5% to 15.3% after the bans; in graduate engineering programs, it dropped from 6.2% to 4.6%; and in the natural sciences, it fell from 7.8% to 6.3%.

The negative consequences also extend beyond admissions, making it more difficult for administrators to talk about race and racism, and undermining actions that support the educational experiences of students already enrolled.

The health of the nation and the rigor of its scientific inquiry require postsecondary administrators and leaders to respond to recent court actions and other legal intimidation by the Trump administration with a reinforced commitment to upholding student diversity. This commitment plays a critical role in preserving democracy and bolstering the global economy. Sadly, today's political climate has made efforts to realize these goals extremely difficult.

—Liliana M. Garces



**Liliana M. Garces**  
is an associate professor at the University of Texas at Austin and an affiliate faculty member at the University of Texas School of Law, Austin, TX, USA.  
liliana\_garces@utexas.edu



**“...negative consequences  
also extend  
beyond admissions...”**

“Federal scientists are doing the best they can, and they are remarkably resilient.”

**Jacob Carter of the Union of Concerned Scientists**, on its new survey in which scientists at 16 agencies describe working conditions under President Donald Trump's administration.

## IN BRIEF

Edited by **Jeffrey Brainard**

### NATIONAL SECURITY

## Psychologists affirm detainee ban



Military psychologists are barred from working with detained terrorism suspects.

**T**he American Psychological Association (APA) in Washington, D.C., voted last week to retain a policy banning military psychologists from working with detainees at Guantanamo Bay, the U.S. naval base in Cuba, and other national security detention facilities. But the 115,000-member organization is still at odds about the ethical rules of conduct for U.S. psychologists in the government's ongoing war against terrorism. APA's ruling body rejected by a nearly two-to-one margin a proposal to let military psychologists provide mental health services to detainees who requested help; the proposal retained a ban on participating in interrogations. A request to create a task force to seek a compromise was also rejected. The debate, held on the eve of the group's annual meeting in San Francisco, California, took place behind closed doors after some advocates of lifting the ban said they were being intimidated by opponents.

## Two Salk lawsuits settled

**WORKPLACE** | The Salk Institute for Biological Studies in San Diego, California, announced last week that it has settled out of court with two of three female professors who sued it for gender discrimination last summer. In recent weeks, “productive conversations have led to a resolution of all claims between these parties that will enable us to put our disagreements behind us and move forward together at Salk,” the institute's president, Rusty Gage, wrote in a joint statement with the two former plaintiffs, Kathy Jones, 63, and Victoria Lundblad, 65. Terms of the settlements were not disclosed. Both women will continue as full professors at the storied research center, said their lawyer, Deborah Dixon of Gomez Trial Attorneys in San Diego. A third plaintiff, Beverly Emerson, 66, will continue to press her lawsuit, which is set to go to trial in December. Emerson left Salk after her contract was not renewed in December 2017.

## Court rejects EPA on insecticide

**PUBLIC HEALTH** | A U.S. federal appeals court last week ordered the Environmental Protection Agency (EPA) to prohibit a widely used insecticide linked to health problems in children, overruling a decision by President Donald Trump's administration to reject the ban. In March 2017, then-EPA Administrator Scott Pruitt rejected a petition to ban the pesticide, chlorpyrifos. His move reversed an EPA decision, made under then-President Barack Obama in 2015, to institute a ban. Health studies have shown that food residues containing chlorpyrifos were associated with neurodevelopmental disabilities in children.

## Japan med school favors men

**HIGHER EDUCATION** | Tokyo Medical University (TMU) confirmed last week that for at least a decade, administrators slashed points from the entrance test scores of female applicants to ensure that entering students—and, thus, graduates—were overwhelmingly male. University officials justified the policy by claiming women leave their careers after having children,



which would cause staffing shortages at TMU's affiliated hospitals—a rationale that has infuriated Japanese women, educators, and even politicians. For the academic year beginning in April, 8.8% of male applicants were admitted, but the figure for female applicants was only 2.9%, leaving women making up only 18% of this year's incoming class. The finding also raised suspicions about similar practices at other medical schools. Women made up only 20.3% of Japan's practicing doctors in 2015, according to Organization for Economic Co-operation and Development (OECD) data. That ratio is the lowest among OECD's 35 member countries, and less than half the average. The education ministry has promised to investigate admission procedures at all of the nation's 80 medical schools.

## Gene-silencing drug approved

**DRUG DEVELOPMENT** | The U.S. Food and Drug Administration last week issued its first approval for a drug that relies on RNA interference (RNAi)—a method of silencing disease-causing genes using small pieces of RNA. The technique won its discoverers a 2006 Nobel Prize, but has faced a rocky path during clinical trials because of side effects and the challenge of protecting RNA fragments from degradation in the body. The new drug, sold as Onpattro and developed by Cambridge, Massachusetts-based Alnylam Pharmaceuticals, is approved to treat nerve degeneration caused by hereditary transthyretin amyloidosis, a rare disease in which a misfolded liver protein builds up in the body and damages nerve and heart tissue. Clinical trials revealed only mild to moderate adverse events in people taking Onpattro versus the placebo, including

swelling of the limbs and reactions at the site of infusion. Other RNAi-based treatments for harder-to-target tissues outside the liver are still under development.

## NIH harassment policy questioned

**WORKPLACE** | Two Democratic lawmakers have criticized the U.S. National Institutes of Health (NIH) for what they call an unacceptably lax policy on grant recipients who commit sexual harassment. In a 6 August letter to NIH, Senator Patty Murray (D-WA) and Representative Rosa DeLauro (D-CT) note that its policy merely requires that grantee institutions certify they will comply with federal civil rights laws. NIH also requires institutions to report principal investigators (PIs) they put on administrative leave for any reason. In contrast, the National Science Foundation (NSF) will soon adopt a specific harassment policy; it would require institutions to report NSF-funded PIs found to have committed harassment and those put on administrative leave during an investigation. "[NIH] has largely failed to take steps to hold its awardee institutions accountable for fostering safe workplace environments regarding extramural harassment," the lawmakers write. They requested details on NIH's policies.

## YouTube clarifies science videos

**SCIENCE COMMUNICATION** | As part of its growing effort to fight disinformation, Google's YouTube has begun providing short blurbs offering science-based information, linked to longer explanations on Wikipedia, with its videos about global warming and the measles, mumps, and rubella vaccine. The annotations appear

below algorithmically selected videos that peddle climate and vaccine misinformation and also on fact-based series; they are part of a YouTube campaign targeting topics that have drawn conspiracy theorists, such as the Apollo moon landing. The blurb on climate videos says: "Multiple lines of scientific evidence show that the climate system is warming," although it omits the significant fact that human greenhouse gas emissions have been the primary driver of that warming.

## Repairs begin for damaged scopes

**ASTRONOMY** | The U.S. National Science Foundation (NSF) has provided \$16.3 million for repairs now underway to two radio telescopes damaged by Hurricane Maria last September. Of that total, \$2 million is for repairing a 240-ton radio telescope on St. Croix in the U.S. Virgin Islands that sustained wind and water damage to its electronics and supporting infrastructure; work began last week and is expected to be completed in 2 years. NSF also gave \$14.3 million to repair the Arecibo Observatory, a giant radio telescope in Puerto Rico. The project began in May and is expected to take several months. The St. Croix station is the southernmost of 10 telescopes in an interferometry array stretching across North America that sharpens images of cosmic objects by combining the telescopes' signals. The array's observations continue with nine stations, but the lack of data from St. Croix has degraded some images, and several experiments have been halted.

**S SCIENCEMAG.ORG/NEWS**  
Read more news from Science online.



Many farmers spray glyphosate, sold under the brand name Roundup, to kill weeds.

## LEGAL AFFAIRS

### Jury says weed killer caused cancer

**A**gricultural giant Monsanto must pay \$289 million for failing to provide proper warnings of the alleged carcinogenic effects of glyphosate, the active ingredient in its widely used weed killer Roundup, a jury in San Francisco, California, decided last week. The civil verdict is the first of its kind; Monsanto faces hundreds of additional, similar lawsuits in other U.S. courts from plaintiffs who, like the one in the San Francisco case, allege they developed non-Hodgkin lymphoma after exposure to the chemical. Controversy over glyphosate intensified after the World Health Organization's cancer research agency in 2015 judged glyphosate "probably carcinogenic to humans." But after reviewing safety studies, the U.S. Environmental Protection Agency and European regulators later deemed the product safe if properly used. Last week's civil lawsuit turned on whether Monsanto should have labeled the product's potential risks more clearly. Monsanto said it plans to appeal.



## SCIENTIFIC COMMUNITY

# New case of alleged bullying rocks the Max Planck Society

Colleagues say empathy researcher Tania Singer created an atmosphere of fear at her lab

By Kai Kupferschmidt, in Leipzig, Germany

**O**n 1 August, Martin Stratmann, president of Germany's Max Planck Society (MPG) in Munich, emailed more than 300 MPG institute directors about a growing concern at Germany's basic research powerhouse. "Recent media reports on bullying and harassment at our Institutes are harming [MPG's] reputation," wrote Stratmann, who announced a special task force to address the issue. "We want to know if the events that have taken place at a few institutes are isolated cases or if we are dealing with structural problems."

At the time, only one such case was widely known: Guinevere Kauffmann, director of the Max Planck Institute for Astrophysics in Garching, stands accused of insulting and bullying students and making racist remarks. (Kauffmann has received coaching and monitoring, and now leads a drastically reduced group.) But now, another long-festering case has exploded into full view. Numerous former or current co-workers have accused celebrated neuroscientist Tania Singer, director

of the Max Planck Institute for Human Cognitive and Brain Sciences here—and one of the world's foremost experts on empathy—of bullying and intimidating them.

The researchers, all but one of whom insisted on remaining anonymous because they feared for their careers, describe a group gripped by terror of their boss. Many say Singer often had especially harsh words for women who became pregnant while working at the lab. Singer, who is now on a 1-year sabbatical, declined to answer questions for this article, but has acknowledged making mistakes in the past: "Problems associated to my exhaustion due to having to carry and be responsible for [a] huge and complex study" were partly to blame, she wrote in a 12 February 2017 email to representatives of her department. But in a letter to *Science*, Singer's lawyer strongly denies the bullying allegations.

Coming shortly after the reports about the institute in Garching, published by *Der Spiegel* and BuzzFeed News Germany, Singer's case has raised fresh questions about the way MPG's 84 independent insti-

Tania Singer wants to show that meditation can make people more kind and caring.

tutes operate. MPG, which has an annual budget of €1.8 billion, invests heavily in recruiting top researchers from around the world, then gives them an unusual amount of freedom to run their labs. Some have suggested that creates the circumstances in which bullying can easily occur.

PhDnet, a network of Ph.D. students within MPG, said it would release a statement after *Science* went to press this week to propose ways to protect junior MPG scientists. The group will demand mandatory and regular leadership training sessions for anyone responsible for training early career researchers, for instance, and advisory committees to oversee Ph.D. theses, instead of just one senior scientist. Haakon Engen, a former lab member in Singer's group who now works at the University of Mainz, says he didn't have such a committee: "That's a really difficult situation to be in as it completely isolates you and puts you at the mercy of your [supervisor]."

An MPG spokesperson said the society could not answer questions about the scandals because both Stratmann and Vice President Bill Hansson are on holiday this week. But in a 7 August statement, MPG acknowledged it has been trying to address the Singer case for more than a year. And in a 14 July interview with German newspaper *Frankfurter Allgemeine Zeitung*, Stratmann said the Garching affair had shown that the society's procedures for dealing with complaints did not work well. "I have to concede that, and for this reason we will improve it," he said.

Singer, the daughter of celebrated neuroscientist Wolf Singer, helped found a new field called social neuroscience. She rose to prominence with her work on empathy, including a landmark study published in *Science* (20 February 2004, p. 1157) that showed watching a loved one experience pain activates the same brain areas as feeling physical pain directly. In 2013, she started an ambitious study, The ReSource Project, in which 160 participants were trained for 9 months to demonstrate that meditation can make people more kind and caring. "She's creative, she can be charming, she knows how to make contacts and get resources. It's a gift and it was necessary to make a project like this happen," one colleague says. "Her superpower is vision," another adds.

But colleagues say working with Singer was always difficult. She wanted to control every detail but was often not available for discussion, they say, and in-person meetings could quickly turn into a nightmare. "She gets extremely emotional and when that



turns dark it is terrifying,” one colleague says. “Whenever anyone had a meeting with her there was at least an even chance they would come out in tears.”

Almost every current or former lab member brought up Singer’s treatment of pregnant women. “People were terrified. They were really, really afraid of telling her about their pregnancies,” one former colleague says. Bethany Kok, a former lab member, says Singer reacted kindly when she first told her she was pregnant with twins. But the next day, Kok says, “She started screaming at me how she wasn’t running a charity, how I was a slacker and that I was going to work twice as hard for the time I would be gone.” A few weeks later, Kok says, she miscarried one of the twins and missed a lab meeting for an urgent medical appointment; she says Singer reproached her in an email.

The issue was also on a list of grievances, shared with *Science*, that lab members say they brought up in a meeting with the institute’s scientific advisory board in February 2017 and recorded afterward. “Pregnancy and parental leave are received badly and denied/turned into accusations,” the notes say. Other complaints aired at the meeting, according to lab members’ notes, included “emotional abuse, threats, devaluation of work, and personal abilities.”

Singer’s lawyer wrote to *Science* that Singer never discriminated against pregnant women or any other group, and said the events described by others “either did not happen or they happened very differently than described.” The letter said Singer had “apologized deeply” during a 2017 mediation process and had taken responsibility for the problems, for instance by asking for the sabbatical and for the appointment of a temporary replacement. It also suggested that the allegations against her came from a “subgroup with its own strong interests and group dynamics.”

In its 7 August statement, MPG said Hansson had investigated the allegations, but that details are confidential. MPG said that “to calm the situation down,” it agreed to Singer’s sabbatical, which took effect in January. In a plan presented to the researchers on 25 July, MPG said it would separate Singer from her current colleagues and allow her to set up a new, smaller research group in Berlin for 2 to 3 years while the postdocs and Ph.D. students here finish their projects and move on; Singer would then return to her lab. (The Leipzig group, which once numbered more than

20 scientists, has dwindled to just five.) The plan would take effect in January 2019, according to official minutes from the 25 July meeting, obtained by *Science*.

Singer “has learned the lesson that groups that are too big carry the risk of losing contact with co-workers,” her lawyer writes. “A good work environment and dealing with each other respectfully are important to her.” But Singer’s colleagues say they are speaking out now to prevent her from regaining her management position. “It appears the Max Planck Society decided it would rather sacrifice another generation of students than risk a scandal,” says one former colleague. Asked how MPG would ensure that future students are treated better, a spokesperson says details of the plan are still being discussed.

To investigate and adjudicate future conflicts at MPG labs, the PhdNet paper advocates establishing an independent committee that would include early career scientists and would be headed by a professional mediator, not an MPG employee. “Control over the arbitration process, flow of information and communication has to always lie with the victim,” the doctoral students say. “The president is open to a dialogue with PhdNet on this matter,” the MPG spokesperson says.

One problem is that MPG directors often are very isolated, says biochemist and Nobel laureate Thomas Südhof of Stanford University in Palo Alto, California, who held an MPG director’s post in the 1990s. “They don’t really have peers around them that can criticize them,” he says. “If I treated postdocs badly here in Stanford it would quickly be known, because I have a hundred colleagues whose postdocs talk to my postdocs.” But the MPG spokesperson argues that problems are rare among the society’s 309 institute directors. “Two cases are less than 1%, you can hardly deduce a structural problem from that.”

Engen, the former member of Singer’s lab, says he used to think “there was something horribly wrong with the MPG in particular” but isn’t so sure today; MPG might just attract more media attention than universities with similar problems, he wrote in an email. The underlying problem is that good scientists aren’t necessarily good leaders, Engen added. “What is critical is that there are safety valves and procedures that enable meaningful oversight over the people who are in charge, and that the people paying the bill (i.e. the politicians) engage with the problem.” ■

## EVIDENCE-BASED POLICY

# Critics pan EPA plan for weighing toxic chemical risks

Guidance could allow regulators to exclude key studies, researchers argue

By Vanessa Zainzinger

**A**cademic scientists and advocacy groups are urging the Environmental Protection Agency (EPA) to withdraw and rewrite proposed guidelines for determining which scientific findings to use when evaluating the safety of toxic chemicals. Critics say that, if adopted, the guidance will allow regulators to exclude high-quality health and risk studies for “ridiculous” reasons, favor industry-backed research, and prevent EPA from considering academic studies that rest on innovative methods.

EPA’s guidance “is less about evaluating the quality of evidence, and more about eliminating it altogether,” the Natural Resources Defense Council (NRDC) of Washington, D.C., wrote this week to EPA in comments blasting the “flawed” proposal, which it says “describes a head-in-sand approach to any evidence that a toxic chemical is toxic.” EPA, however, says the guidelines are likely to evolve and that it is aiming for an “efficient systematic review process that generates high-quality, fit-for-purpose risk evaluations that rely on the best available science.”

The controversy, which mirrors a debate over a proposal EPA released earlier this year that critics say would allow the agency to ignore certain human health studies, has its roots in a 2016 overhaul of the nation’s premier chemical safety law, the Toxic Substances Control Act (TSCA). The revised law aims to make it easier for EPA to complete safety reviews of new chemicals before they reach the market and to more quickly restrict the use of existing chemicals if new evidence of risks emerges. It also orders EPA to develop new guidelines for the “systematic review” of the quality of the scientific evidence used in risk assessments.

In response, EPA released a 248-page draft in May describing the review process

***“We want to know if the events ... are isolated cases or if we are dealing with structural problems.”***

**Martin Stratmann,**  
Max Planck Society

it is using for its first major round of chemical assessments under the new TSCA. The agency is focusing on 10 compounds already in use, including asbestos (which is still used in the production of chlorine), carbon tetrachloride (often used to make refrigerants), and Violet 29, a common pigment. EPA notes the document offers only “general expectations” and that regulators can depart from the guidance. But it does appear to allow a publication to be excluded if it does not include detailed methods, for instance, or if specific information about study subjects or data sources is withheld because of privacy, business, or other concerns.

A key industry group, the American Chemistry Council in Washington, D.C., said in a statement that the guidelines have “many positive attributes,” but would “benefit from additional explicit guidance,” for example on how to make EPA’s process for integrating evidence into the rulemaking process more transparent.

Critics worry that President Donald Trump’s appointees at EPA will use the guidance—which they note has not been independently reviewed—to put a pro-industry spin on evidence reviews, and perhaps make it easier for companies to keep dangerous chemicals on the market. And they

say it could make it easier for EPA to ignore data from certain kinds of health studies. “The things that they are excluding [studies] on are ridiculous,” says Tracey Woodruff, director of the Program on Reproductive Health and the Environment at the University of California, San Francisco (UCSF), one of 25 academics and advocacy group officials who signed a letter to EPA criticizing the guidance. In particular, they say it places too much emphasis on how a study is described—or reported—in the scientific literature.

Detailed information on methods and subjects is often left out of journal publications, the critics note, and has “nothing to do with the quality of the study,” Woodruff says. The approach, NRDC adds, mirrors the so-called transparency rule proposed this past April by former EPA Administrator Scott Pruitt; researchers argue it would bar the agency from using major epidemiological studies based on confidential health records (*Science*, 4 May, p. 472).

In contrast, the critics note, the TSCA guidance fails to mention financial conflicts of interest as a reason for excluding a study. And they argue it inappropriately implies that studies done in industry-funded laboratories that adhere to specific standards, known as Good Laboratory Practices (GLP), are more trustworthy than studies done in academic laboratories that don’t follow GLP—often because they are pioneering innovative methods (*Science*, 10 February 2017, p. 564).

Critics want EPA to scrap the draft guidance and replace it with an established method for systematic review, such as one developed by the National Toxicology Program, which has already been peer reviewed. “I have not seen an explanation of why EPA’s approach is so far outside of the scientific mainstream on systematic reviews,” says Veena Singla, an associate director of UCSF’s health and environment program.

EPA is now deciding how to proceed. The deadline for public comment on the draft was 16 August. The guidelines would need to be revised soon, as the agency has said it wants to complete its 10 initial chemical assessments by late next year. ■

*Vanessa Zainzinger is a journalist in the United Kingdom.*

## ARTIFICIAL INTELLIGENCE

## AI takes on video games in quest for common sense

Bots seek to hone strategy skills in matchup with top *Dota 2* players

By **Matthew Hutson**

**N**ext week, scientists working on artificial intelligence (AI) and games will be watching the latest human-machine matchup. But instead of a single pensive player squaring off against a computer, a team of five top video game players will be furiously casting magic spells and lobbing (virtual) fireballs at a team of bots called OpenAI Five. They’ll be playing the real-time strategy game *Dota 2* at The International in Vancouver, Canada, an annual e-sports tournament that draws professional gamers who compete for millions of dollars.

In 1997, IBM’s Deep Blue AI bested chess champion Garry Kasparov. In 2016, DeepMind’s AlphaGo AI beat Lee Sedol, a world master, at the traditional Chinese board game Go. Computers have also defeated humans in checkers and some forms of poker. But fast-paced multiplayer video games pose a different kind of challenge, requiring computers to collaborate and manage unpredictability. The goal is common sense, which could help AIs handle real-world situations such as navigating traffic and providing home care—even if they never have to face a magic spell.

“The next big thing for AI is collaboration,” says Jun Wang, a computer scientist at University College London who works on *StarCraft II*, another real-time strategy game. That requires “strategic reasoning, where it’s understanding the incentives of others,” says Jakob Foerster, a computer scientist at the University of Oxford in the United Kingdom, who also works on *StarCraft II*.

*Dota 2*, released in 2013, has millions of players around the world. In a game, teams fight to destroy a structure on their enemy’s turf while defending their own, all the while collecting resources to increase their



U.S. regulators are reviewing the safety of asbestos, which is still used to make fire-resistant fabrics.





In warmup matches, OpenAI Five defeated human players at the multiplayer strategy game *Dota 2*.

strength and skills. A well-matched game lasts about 45 minutes. A year ago, OpenAI, a research nonprofit based in San Francisco, California, revealed an AI that could beat the best human players in one-on-one games. But the five-on-five matchups showcased at The International present a much bigger challenge for a computer because the games are longer and more chaotic, says Greg Brockman, OpenAI's co-founder and chief technology officer. Still, in a warmup exhibition last week, OpenAI Five easily beat a team of former pro players. "It sucks getting embarrassed by a nonperson," says William "Blitz" Lee, who lost on stage in front of a live audience. "We were just getting crushed left and right."

***"It sucks getting embarrassed by a nonperson. We were just getting crushed left and right."***

**William "Blitz" Lee,**  
*Dota 2 player*

The range of possible moves in *Dota 2* is far greater than in chess or Go, where each move has at most a few hundred options. In *Dota 2*, the action is constant, and players have thousands of options per move—where to flee, which spell to use, where to aim it. Such freedom, combined with the game's inherent randomness and players' ignorance about what's out of view, means you can't perfectly predict what the game will look like even one move ahead. In chess and Go, algorithms use search trees, analyzing branching possibilities far into the future. In *Dota 2*, forecasts become fuzzy much more quickly.

So instead of relying on search trees, OpenAI Five uses neural networks, algorithms inspired by the brain that strengthen connections between small computing elements in response to feedback. (AlphaGo combined neural nets with search trees.) During training, the system blindly experiments with different moves in the game. When they perform well, the connections responsible for those acts are reinforced. After thousands of years of (sped-up) gameplay, strong strategies emerge. OpenAI applied this method, known as reinforcement learning, on a massive scale, running the algorithm on thousands of computers at once. "OpenAI Five is one of the most impressive demonstrations of reinforcement

learning I have seen," says Niels Justesen, a computer scientist at the IT University of Copenhagen who also works on *StarCraft II*.

Parsing last week's warmup performance, Michael Cook, a computer scientist at Falmouth University in the United Kingdom who studies AI and games, says OpenAI Five excels by relying on a "superhuman ability to calculate the outcome of certain actions," such as the damage a particular attack will inflict on an opponent. OpenAI handicapped it to have the same reaction time as a human player, about a fifth of a second, but in that blink the system processes much

more information. Such thoroughness and precision make it deadly at fights, leading to "a Blitzkrieg-like approach to the game," Cook says. "It's amazing to watch." But the aggressiveness may mask a weakness in long-term strategizing, Cook suggests: In one game that it lost, the AI was assigned characters that needed more time to build up their abilities, and it couldn't adjust.

On the face of it, OpenAI Five also appears to succeed at collaboration. The AI's five players were quite willing to be killed for the overall good of the team, which might confer an advantage over the human teams. "The bot plays very sacrificially," Lee says. Humans are less likely to give up a player in order to win, he says. "It's a very human concept to be greedier." But the AI relies on a kind of hive mind that may make coordination easier. Each of five nearly identical algorithms in the system gets a peek at what the others see, whereas humans see only what's on their own screens and share information only by talking. To collaborate with people or programs unlike themselves, whether in games or in life, Wang says the algorithms will eventually need to develop communication skills and "theory of mind"—models of the beliefs and desires of other people and algorithms.

*StarCraft II*, the game many AI researchers prefer to work on, may be better for honing long-term planning. It is more like economics, Wang says, as it entails managing a colony's resources as it builds weapons factories. And other games may provide better tests of an AI's ability to model the mind of an opponent than *Dota 2*; Foerster mentions strategy board games such as Settlers of Catan and Risk, where multiple players must negotiate, trade, and form alliances, both cooperating and competing.

Even so, *Dota 2* remains a worthy test for an AI. Many experts are expecting OpenAI Five to win at The International. But Vanessa Volz, a computer scientist at the Technical University of Dortmund in Germany who studies AI and games, sees a potential weakness, in that OpenAI Five uses self-play to train its algorithms. "This approach has the risk of being vulnerable to previously unseen playing styles," she says. Lee, who lost to the AI, feels the same way. "Right now, the bot is a little too rigid," he says. "It's starting to get a little too predictable. I feel like if we'd had a few more games we would have been able to take games off it pretty cleanly." ■

*Matthew Hutson is a journalist in New York City.*



Thera, now called Santorini, erupted sometime in the Bronze Age, creating a caldera now mostly filled with water.

## ARCHAEOLOGY

# Study reignites debate about when Thera blew its top

Radiocarbon curve suggests archaeological data were right

By Lizzie Wade

**H**undreds of years before the Trojan War, the volcanic island of Thera in the Aegean Sea blew its top in an explosion that rocked the ancient world. Sixty times greater than the 1980 eruption of Mount St. Helens in Washington, the blast completely buried the Theran town of Akrotiri and sent 12-meter-high tsunamis hurtling toward the heart of the Minoan civilization on Crete, 110 kilometers to the south. Some authors have even speculated that the Atlantis myth may stem from a cultural memory of the cataclysm.

But when exactly did it happen? The eruption spread ash across the eastern Mediterranean, so a precise date could pin down the chronologies of ancient cultures including the Greeks, Minoans, and Egyptians. Archaeologists and radiocarbon daters have battled fiercely over the timing. By correlating Egyptian records and pottery, archaeologists put the eruption as early as 1500 B.C.E. But radiocarbon dates from Akrotiri and nearby sites, including an olive tree buried by the eruption, pointed to a date more than 100 years earlier, in the late 17th century B.C.E. (*Science*, 28 April 2006, p. 508).

A new study promises a truce. A team led by archaeologist and tree ring scientist Charlotte Pearson of The University of Arizona in Tucson measured the radiocarbon stored in individual rings from five trees. Because tree rings can also be dated by simply counting them, the team could correct the radiocarbon dates. That broadens the possible dates of the eruption to include the traditional archaeological date of the 16th century B.C.E.

"It's a very impressive data set," says Paula Reimer, a geochronologist at Queen's University Belfast and chair of the International Calibration Working Group (IntCal), which establishes the worldwide radiocarbon calibration curve. If another lab can reproduce the findings, she says, every radiocarbon date between 1700 B.C.E. and 1500 B.C.E. might need to be recalibrated.

The radiocarbon clock doesn't tick steadily. Organisms contain both carbon-14 ( $^{14}\text{C}$ ) and carbon-12 ( $^{12}\text{C}$ ). At death,  $^{14}\text{C}$  starts to decay at a known rate while  $^{12}\text{C}$  levels stay the same. By comparing the ratio of isotopes, scientists can calculate how long ago an organism was alive. But  $^{14}\text{C}$  in the atmosphere, and therefore in all organisms, fluctuates with the amount of cosmic radiation hitting Earth. To calibrate the clock, scientists must track those fluctuations. That's where tree rings come in, because they yield both a radiocarbon reading and an absolute age.

The IntCal curve was largely built using 10-year chunks of wood. But Pearson's team was able to measure the radiocarbon more precisely, in single, annual rings of

three ancient bristlecone pines in California and two oak trees in Ireland. The resulting calibration curve differs slightly but significantly from the IntCal curve, with a so-called radiocarbon plateau that allows a much broader range of eruption dates. Some fall in the late 16th century B.C.E., perhaps around 1540 B.C.E.—closer to the archaeological date.

"This slight adjustment to the shape of the calibration curve can allow all the lines of evidence to overlap," Pearson says. "It doesn't give us the definite date but it moves us a step in the right direction."

For radiocarbon experts, the result is bitter-sweet, vindicating archaeologists without challenging the radiocarbon measurements themselves. Christopher Ramsey of the University of Oxford in the United Kingdom helped date the Thera eruption to the 17th century B.C.E. and has fielded years of critiques of his dates; just last week, for example, a paper in *Scientific Reports* showed that olive trees grow irregularly and might give artificially old dates. But if Pearson's paper is correct, Ramsey says, "then all the measurements are fine!" It's the calibration data—the gold standard used by researchers around the world—that may be off. "That was not something I expected," he says ruefully.

Sturt Manning, an archaeologist at Cornell University who led the earlier radiocarbon work, worries the new findings will encourage traditionalists to disregard radiocarbon evidence altogether. "This paper will now be cited for the next decade to demonstrate that there's ambiguity with the radiocarbon, and therefore we should ignore it," he says. He emphasizes that the new calibration data must be confirmed independently.

Jeremy Rutter, an archaeologist emeritus at Dartmouth College, considers the work "a real step forward." But, he says, "I'm a little disappointed by where we end up," with no absolute date for the eruption.

"It's very bad luck that this really important eruption happens to occur on a radiocarbon plateau," Pearson says. "But on the other hand, this is science, isn't it?" ■





## AGRICULTURE

# Detailed genome maps paths to better wheat

Chromosome-by-chromosome sequence helps navigate crop's genomic complexity

By Elizabeth Pennisi

**T**he world's most widely planted cereal has also proved to be among the hardest to improve. Plant breeders vastly increased wheat yields during the Green Revolution of the 1960s, but since then efforts to improve the crop through traditional breeding or genetic technology have been painstakingly slow because of the fiendish complexity of its genome. Thanks to a decadelong effort, the wheat genome has finally come into sharp focus, speeding the search for genes that could boost harvests and even make wheat less likely to trigger allergies.

The data, described on pp. 661 and 662, represent the long-awaited culmination of the International Wheat Genome Sequencing Consortium, a massive collaboration of academic and industry researchers from 20 countries. Wheat geneticists say the newly finished genome, which pinpoints 107,000 genes on the 21 chromosomes of bread wheat, has transformed their research as they and others gained early access to it. "What took us years in the past now takes us one night," says Jorge Dubcovsky of the University of California, Davis, who recently found a new gene for wheat height. "It's like walking with a Google map."

Already the new genome has helped plant geneticists from the John Innes Centre (JIC) in Norwich, U.K., to boost grain size by 20% in lab-grown wheat. In a preprint posted on the bioRxiv server in May, they report identifying multiple copies of a gene for grain size originally found in rice, then bulking up wheat grains by mutating the genes using CRISPR gene-editing technology. Many more traits beckon. The new sequence "ushers in a new era in wheat genetics," says James Anderson, a plant breeder at the University of Minnesota in St. Paul who was not part of the effort.

For him and thousands of other wheat researchers, bread wheat's DNA has been an impenetrable thicket. Natural breeding between two grasses many thousands of years ago gave rise to the durum wheat now used in pasta. That hybrid was mated with yet

another grass to yield the grain that makes everything from bread to beer all over the world. But this interspecies mingling produced a genome more than five times the size of the human one, harboring three sets of very similar chromosomes, 21 pairs in all, with six copies of most genes. Lacking a map of this complex landscape, wheat breeders had trouble tracking genes from generation to generation to see whether a crossing had worked, and genetic engineers trying to alter a specific DNA sequence often did not know where to find it.

The consortium—born in 2005 as an initiative by Kansas farmers—has finally mapped the thicket by breaking out and

active under conditions including drought, pest attack, and other kinds of stress, aiming to trace networks of genes underlying yield and other traits. With these surveys, "we can see which of the copies are most useful to target," Borrill says.

So far, consortium members and others have drawn on the genome for more than 100 published papers, says the consortium's executive director, Kellye Eversole, who is based in Bethesda, Maryland. Their finds are already being put to use. For example, commonly planted varieties won't sprout unless the seeds have overwintered in the ground. Last year, plant geneticist Antje Rohde, now at BASF in Ghent, Belgium, reported her team had pinned down a key gene responsible for the sprouting delay. By disabling that gene using CRISPR, the team hopes to shorten the wheat breeding cycle.

The genome could also help bolster wheat's resistance to disease. Drawing on the consortium data, Curtis Pozniak and Kirby Nilsen at the University of Saskatchewan in Saskatoon, Canada, found a gene that makes wheat stems stiffer, and hence more resistant to stem-boring insect pests called sawflies. Stiffer wheat has more copies of the gene, Nilsen found, which points to ways to protect other wheat varieties.

The genome could even aid human health, says consortium co-leader Rudi Appels, a molecular geneticist at Murdoch University in Perth, Australia. In *Science*

*Advances* this week, he, fellow Murdoch University researcher Angéla Juhász, and Odd-Arne Olsen from the Norwegian University of Life Sciences near Oslo report identifying 365 genes coding for wheat proteins that stimulate an immune or allergic response. The data could help breeders aim for less problematic wheat—what Appels calls "my personal dream."

"For the first time, people working in wheat have the quality of resources that people have in other crops," says consortium co-founder Catherine Feuillet, chief scientific officer at Inari Agriculture, a new ag-biotech startup in Cambridge, Massachusetts. "We now have the tools to do [breeding] in a knowledge-based way." ■



To sprout in the spring, winter wheat must be planted the previous autumn—an extended growing cycle that some breeders would like to shorten.

sequencing each chromosome separately. In contrast, a genome published last year by a group of academics had some longer stretches of contiguous sequence but did not describe genes or order and orient them on the chromosomes. The new genome "represents a major step forward," says JIC's Michael Bevan, whose group also produced a rival draft wheat genome last year, before working with the consortium.

JIC researchers including Cristobal Uauy, Ricardo Ramírez-González, and Philippa Borrill have now built on the consortium's data to document gene activity in different tissues and at various points of the plant's life cycle. They took 850 snapshots of messenger RNA levels to gauge which genes are





**T**he first thing that went through Alison Avenell's head when she heard Yoshihiro Sato had died was that it might be a trick. It was March 2017, and in the previous years, Avenell, a clinical nutritionist at the University of Aberdeen in the United Kingdom, had spent thousands of hours combing through Sato's papers, together with three colleagues in New Zealand. They had discovered that Sato, a bone researcher at a hospital in southern Japan, had fabricated

data for dozens of clinical trials published in international journals. "With so much going on, so much fabrication, you just wonder if it's convenient for the person to go and hide," Avenell says.

Her second thought was that Sato might have killed himself. "We have no indication that he committed suicide, but it concerns us," Avenell said when I met her at her office in late 2017. Three years earlier, Japanese stem cell scientist Yoshiki Sasai had hanged himself in the stairwell of the RIKEN Center for Developmental Biology

in Kobe after he was caught up in a stem cell scandal. "We were aware of the culture in Japan and the dishonor something like this could bring," Avenell said.

It was one more mystery in a deeply unsettling case.

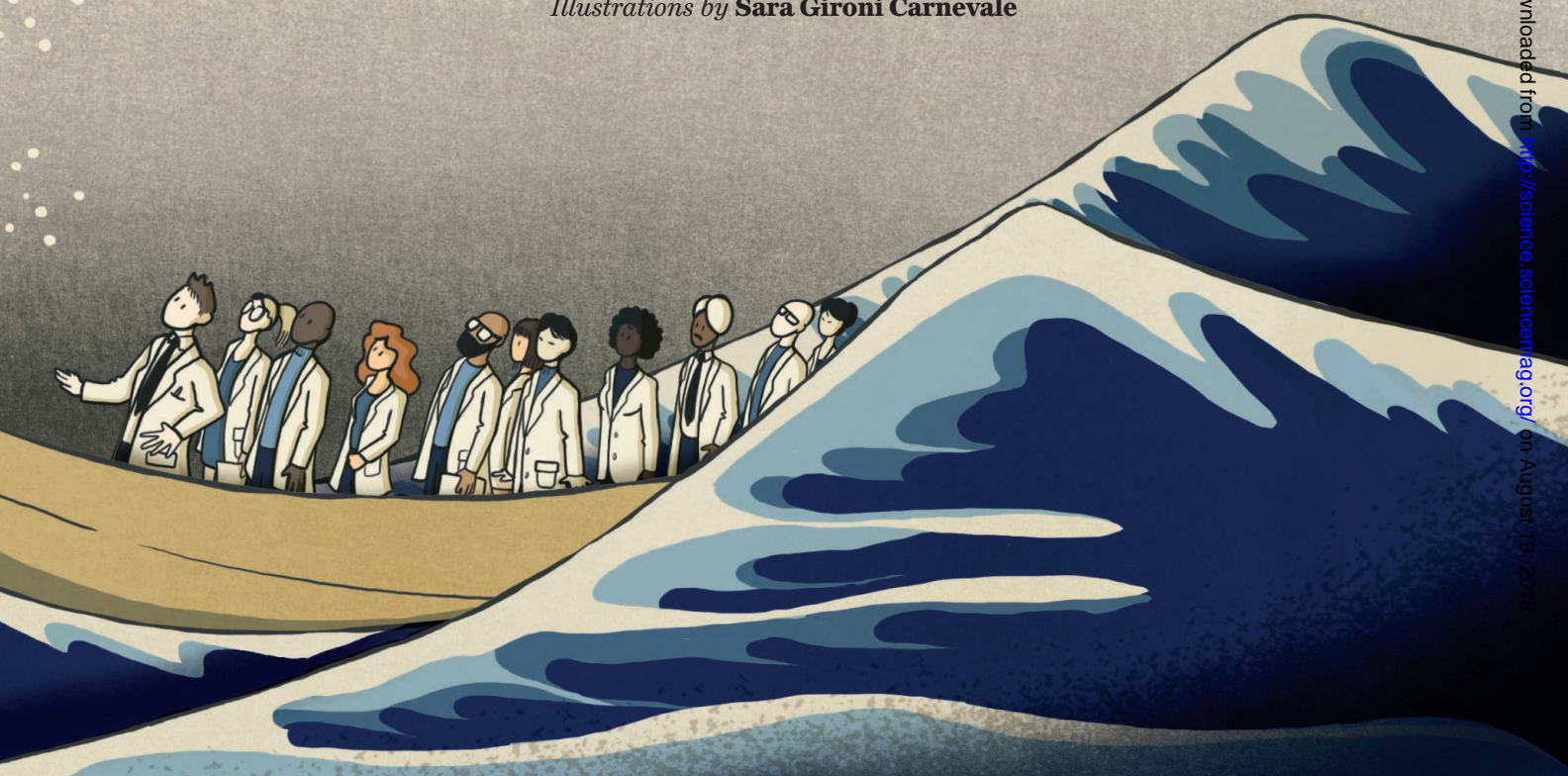
Sato's fraud was one of the biggest in scientific history. The impact of his fabricated reports—many of them on how to reduce the risk of bone fractures—rippled far and wide. Meta-analyses that included his trials came to the wrong conclusion; professional societies based medical guidelines



# TIDE OF LIES

The researcher at the center of an epic scientific fraud remains an enigma to the scientists who exposed him

By **Kai Kupferschmidt**, in Japan;  
Illustrations by **Sara Gironi Carnevale**



on his papers. To follow up on studies they did not know were faked, researchers carried out new trials that enrolled thousands of real patients. Exposing Sato's lies and correcting the literature had been a bruising struggle for Avenell and her colleagues.

Yet they could not understand why Sato faked so many studies, or how he got away with it for so long. They puzzled over the role of his co-authors, some of whom had their names on dozens of his papers. ("Do we honestly believe they knew nothing at all about what was going on?" Avenell asked.) They

wondered whether other doctors at his hospital read Sato's work—and whether the Japanese scientific community ever questioned how he managed to publish more than 200 papers, many of them ambitious studies that would have taken most researchers years to complete.

The tools of science that the group had used—analyzing studies, calculating statistics, writing papers—could reveal fraud. But they could not expose the personal and cultural factors that drove it, or assess its emotional toll. So I set off on a quest that

would eventually lead me to the Mitate Hospital in Tagawa, a small town on the island of Kyushu, where Sato had worked in the last 13 years of his life.

## I: SUSPICION

Avenell's own quest began in 2006, when she was combing through dozens of papers for a review evaluating whether vitamin D reduces the risk of bone fractures. In two papers by Sato, she stumbled on a weird coincidence. They described different trials—one in stroke victims, the other in



Parkinson's disease patients—but the control and study groups in both studies had the exact same mean body mass index. Looking further, she quickly found several other anomalies. She decided not to include Sato's studies in her analysis.

She wasn't the first to notice something was off. In a 2005 *Neurology* paper, Sato claimed that a drug named risedronate reduces the risk of hip fractures in women who have had a stroke by a stunning 86%. In a polite letter to the journal, three researchers from the University of Cambridge in the United Kingdom noted that the study was "potentially of great importance," but marveled that the authors had managed to recruit 374 patients in just 4 months.

Two years later, a letter in what was then the *Archives of Internal Medicine* was less polite. A study of male stroke patients published by Sato had managed to enroll 280 patients in just 2 months; another one, of women with Alzheimer's disease, recruited a staggering 500 in an equally short period. Sato claimed to have diagnosed all of the Alzheimer's patients himself and done follow-up assessments of all 780 patients every 4 weeks for 18 months. Both studies had very few dropouts, and both showed risedronate, again, to be a resounding success. "We are deeply concerned whether the data provided by Sato et al are valid," Jutta Halbekath of *Arznei-Telegramm*, a Berlin-based bulletin about the drug industry, and her co-authors wrote. Sato apologized in a published response and claimed the study had been conducted at three hospitals, not one. "The authors did not describe this fact, the reason being that these hospitals were reluctant to have their names in the article," he wrote. He didn't name the other hospitals or explain why they wanted to remain anonymous. The journal apparently accepted the explanation.

The letter's authors also spotted a troubling pattern. In addition to the two papers in the *Archives of Internal Medicine*, they found 11 further studies by Sato, published elsewhere, that tested whether sunlight, vitamin D, vitamin K, folate, and other drugs could reduce the risk of hip fractures. All but two reported "extremely large effects with significant results," they noted. But the *Archives of Internal Medicine* didn't want to point fingers at other journals. "You may allude to your concern that other papers have similar concerns," its editors warned Halbekath, "but we cannot allow you to mention those other papers by journal name."

By now, several researchers had raised red flags and waved them for everyone to see—and then everybody moved on. "The trail just went cold," Avenell says.

## II: EVIDENCE

Mark Bolland had never heard of Sato when Avenell first mentioned him in late 2012. She and Bolland, a clinical epidemiologist at the University of Auckland in New Zealand, have never met in person, but they joined forces to write meta-analyses on calcium supplements in 2008, together with Andrew Grey and Greg Gamble, both also at the University of Auckland. One topic the quartet discussed frequently was why meta-analyses on the same topic sometimes reach different conclusions. Avenell mentioned Sato's studies and noted that the effects they reported were so strong that they might swing meta-analyses if they were included.

Intrigued, Bolland looked up the papers. He, too, was stunned by the large cohorts, the low number of dropouts, and the big effects of almost any treatment tested. "There is nothing that I can think of that produces a 70% to 80% reduction in hip fractures, yet Sato was able to do it consistently in all his trials," he says.

---

***"We have no indication  
that he committed suicide,  
but it concerns us."***

Alison Avenell, University of Aberdeen

To follow up on his suspicions, Bolland turned to statistics. When scientists compare a treatment and a control group, they usually report "baseline characteristics" for each—things like age, weight, and sex, or, in osteoporosis studies, bone density and calcium intake. From these values, scientists can calculate p-values that are a measure of the similarity of two groups for a given characteristic; the closer to one the value is, the more the groups resemble each other. Because the groups are randomly selected, the p-values should normally be "equally distributed"; the value for age or weight is just as likely to be between 0 and 0.1 as between 0.9 and 1.0, for example.

Bolland extracted the baseline characteristics from the 33 clinical trials Sato had published at the time, more than 500 variables all in all, and calculated their p-values. More than half were above 0.8, he found. "That just shouldn't happen," he says. "The randomized groups were incredibly similar." There was just one plausible explanation, he says: Sato had fabricated data for both groups and had made them more similar than they would ever be in real life.

The team felt it had a damning indictment. "I thought: 'This is so convincing. Everybody is going to believe this,'" Avenell

says. Still, "It needed detailed statistical refereeing, and it needed to be published by a journal so that other affected journals would take note," she adds. So they wrote their accusation as a scientific paper. All they had to do was publish it and wait for researchers, journals, and institutions to react, investigate, and retract. Or so they thought.

## III: ACCUSATION

In March 2013, the team submitted the manuscript to *The Journal of the American Medical Association (JAMA)*, the highest profile journal Sato had published in, and one it felt might have the resources for an in-depth investigation. After reviewing the evidence, *JAMA* Editor-in-Chief Howard Bauchner told the team the editors would ask Sato and, if necessary, his institution to respond.

Two years later, in April 2015, *JAMA* told the researchers the hospital had not responded, and it would publish an "expression of concern"—a short note to flag Sato's *JAMA* paper as suspicious. It would not publish the whistleblowers' paper, however; if the team had concerns about other papers, it should contact the journals that had published them, Bauchner said.

The four researchers were shocked. "To find out after waiting 2 years that in fact nothing much had really happened and, other than an expression of concern, was going to happen in *JAMA*, was quite frustrating," Bolland says. (Bauchner declined to answer *Science's* questions about the case.)

Next, the paper was rejected by *JAMA Internal Medicine*, which had also published Sato's work. The *Journal of Bone and Mineral Research*, a highly rated journal in the osteoporosis field, said it would investigate Sato's papers, but would not publish the manuscript either. The editors of *Trials*, which had not published Sato's work, said it would not be appropriate to get involved.

Bolland became demoralized. The other three persuaded him not to give up. "If you ever embark on something like this, make sure you have a good support team," he says now. Avenell, too, was sometimes despondent. Whereas the other three researchers at least saw each other in Auckland, she was on her own, frustrated, in the dreary, gray town of Aberdeen. Sometimes, she says, she would just sit in a corner of her open floor plan office and cry.

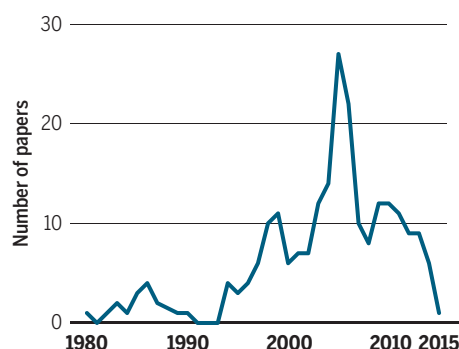
Then, in June 2015, came a small success: *The Journal of Bone and Mineral Research* retracted one of the 33 trials the team had analyzed. A few other journals followed suit in the months after. But some seemed irritated by the group's persistence. "It is apparent that the responses to the *JAMA* inves-



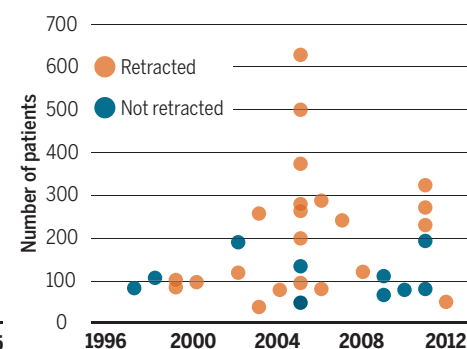
## A far-reaching fraud

A team of four researchers has worked since 2012 to expose scientific misconduct by Japanese bone researcher Yoshihiro Sato, who published more than 200 papers before he died in 2016. The team has focused on Sato's 33 clinical trials, together involving 5894 patients.

### Total scientific output



### Clinical trials



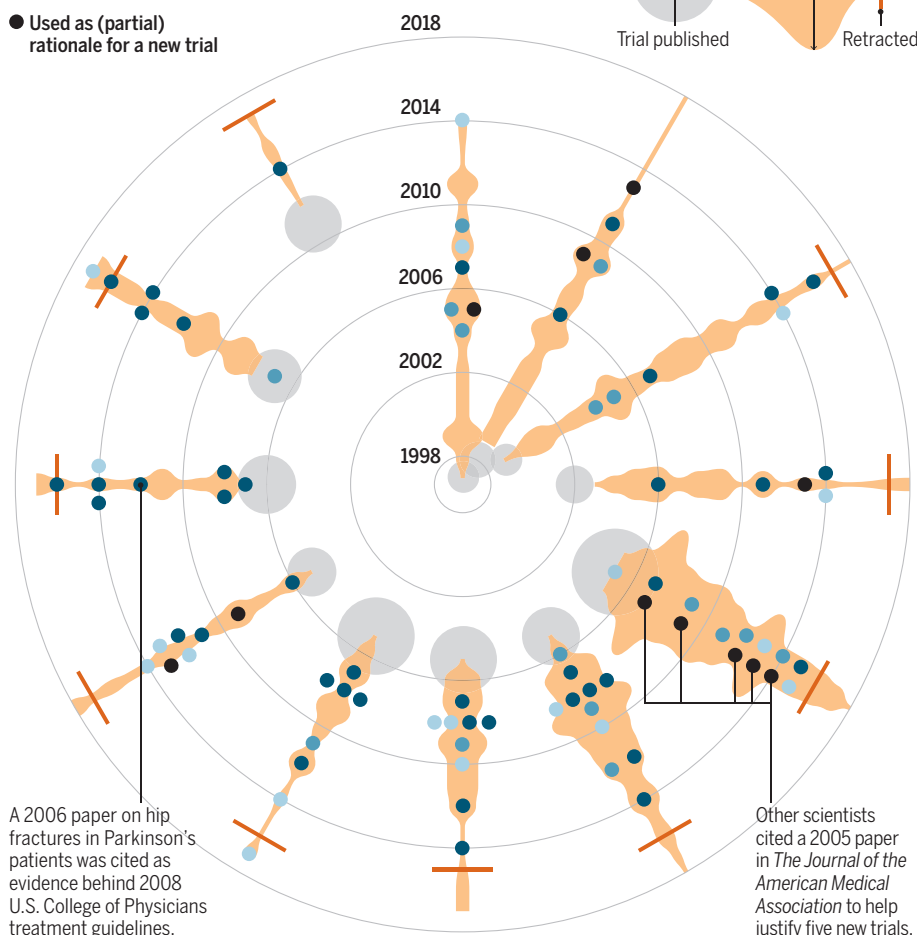
## Ripple effects

The 12 trials Sato published in high-impact journals have been widely cited. Many were included in meta-analyses, sometimes changing the outcomes, or were translated into treatment guidelines. Other researchers used Sato's fake data as part of the rationale for launching new clinical studies.

### Included in systematic reviews, meta-analyses, or treatment guidelines

- Changed the outcome
- Unclear whether outcome changed
- Did not change the outcome

### Used as (partial) rationale for a new trial



tigation by Dr. Sato and his institution have been either inadequate or not forthcoming," Grey wrote to Bauchner in December 2015. "At what point will *JAMA* consider more decisive action, such as retraction?" "We will consider your opinion about how you think it best we should conduct the investigation," Bauchner responded. "We often hear from people how they think we should perform our responsibilities as editors."

In what Bolland calls "really just the last throw of the dice," that same month the group submitted the paper to *Neurology*, where Sato had published three papers about bone fractures in patients with neurological disease. When it was accepted 8 months later, Avenell cried again. "I'm not one usually given to showing such emotion, especially when all I have is a computer screen and emails to look at," she says.

"Journals don't really like going back to investigate when things go wrong," Grey concludes. "They complain that it's time-consuming and laborious and difficult." (It is all of that, says Avenell, "and no one ever thanks you for it.") The group says investigations of this scale should not be handled by journals or institutions; it has suggested a levy on journals to fund an independent investigative body.

By the time *Neurology* published the investigation in December 2016, 10 of the 33 trials had been retracted, all but one by journals the team had contacted. One month later, Avenell received an email from an editor with troubling news. Sato was dead.

## IV: RIPPLES

When scientists die, their published papers live on—even if they're based on lies. Downloaded in seconds from anywhere in the world, fake results continue to steal other scientists' time, influencing their choice of which research avenues to follow and which trials to design and seek ethical approval for.

Today, 21 of Sato's 33 trials have been retracted by the journals or Sato himself; Avenell has crossed them off a list taped next to her computer with a red marker. But now the team is following the ripples that the studies caused, focusing, for the time being, on a dozen papers published in the journals with the highest impact factors. Together, these studies reported results for 3182 participants. They have been referenced more than 1000 times, and 23 systematic reviews or meta-analyses have included one or more of the 12 trials (see graphic, left).

One meta-analysis, which found drugs called bisphosphonates to be highly effective in preventing hip fractures in elderly patients with stroke or Parkinson's, is based

entirely on eight trials from Sato, as he was the only one to study the issue. A key conclusion in another meta-analysis rests only on Sato's four studies on bone mineral density in Alzheimer's patients. Two other meta-analyses would probably come to different conclusions if Sato's trials were removed, Avenell says. One of those, a review showing that vitamin K helps prevent fractures, was the basis of 2011 Japanese guidelines that recommend the supplement for people at risk.

The fake trials led to further, real research. Eight trials referenced at least one of Sato's fabricated papers in explaining the rationale for the trial. Researchers in the Netherlands, for instance, launched a huge study in 2008 to determine whether B vitamins could help prevent hip fractures. Two previous studies found they didn't, but Sato had observed "a large protective effect" in elderly women. "Given the conflicting results and low generalizability to the general older population, further investigation is needed," the Dutch researchers wrote to explain their thinking. The 2-year study in 2919 elderly people found no effect of the vitamins.

The fraud has also drawn attention to the two co-authors whose names appear on Sato's papers most often. One is Kei Satoh, president of Hirosaki University, in a small town at the northern tip of Japan's main island, Honshu. Sato worked at Hirosaki University, where he collaborated with Satoh, until 2003; even after he left for Mitate Hospital, 1600 kilometers to the southwest, he and Satoh remained frequent co-authors, including on 13 of the 33 clinical trials.

Satoh—whose name, confusingly, is sometimes spelled Sato—did not respond to *Science's* emails. In a short letter to Grey, Hirosaki University Vice President Chizuko Kohri wrote last November that the university had asked "three outside experts" to investigate after the *Neurology* paper was published. The committee investigated 38 papers, Kohri wrote. Of these, Sato had already retracted seven and wanted to retract another seven. The committee "concluded that there was research misconduct in these 14 papers," Kohri wrote, but that Sato alone was responsible. According to Japanese press reports, Satoh maintains that he only corrected the English in the papers. As a sign of contrition, he gave up 10% of his salary for 3 months.

Sato's most important collaborator, however, was Jun Iwamoto. A board member of the Osteoporosis Society of Japan, Iwamoto was a senior lecturer at Keio University in Tokyo—one of the country's most prestigious—until 2017, when his contract wasn't renewed in the wake of the Sato affair. He and Sato collaborated for more

than a decade and published more than 130 papers together, including 25 of the 33 clinical trials.

A panel at Keio University has been investigating Iwamoto's clinical trials. Iwamoto told the panel that he first contacted Sato in 1998, when Iwamoto was working at the New York University Winthrop Hospital in Mineola. In 2002 they started to put each other's name on every paper they authored. Still, Iwamoto claims he was unaware of Sato's practice. "We talked to Dr. Iwamoto and in most of the papers

which Dr. Sato published, which included Dr. Iwamoto's name, Dr. Iwamoto did not know that his name was included," says cancer researcher Hideyuki Saya, who heads the investigation. The panel was "very shocked" by this, Saya says. At the same time, he says, "For Dr. Iwamoto it was an honor to put his name on Dr. Sato's [papers] even though he did not know much about the content."

Although considered highly irregular today, such "gift authorships" were common in the recent past, Saya argues. A 2014 study in the *International Journal of Japanese*





*Sociology* found they are particularly common in Japan. “We speculate that most natural science researchers in Japan may be either confused about or struggle with the situation where the strict global criteria conflict with specific local cultures that often condone gift and ghost authorships,” the researchers wrote.

Saya says the seven trials listing Iwamoto as the first author appear not to be fabricated. Data for the first four of those no longer exist, but Iwamoto can’t be faulted for that, says Saya, because under rules at the time they were conducted, he had to save the data for only 5 years. Iwamoto did provide data for three more recent trials. “That data, it seems, was really collected,” Saya says.

But Avenell and her colleagues say they have uncovered many problems in trials on which Iwamoto was a first author as well. For instance, two of them, which tested a drug named alendronate, seem to include the same group of 25 patients, as indicated by their average age, height, serum calcium, and numerous other characteristics, but the two papers give different recruitment dates and inclusion criteria, and some of the outcome data differ. Saya chalks problems in the papers up to “immaturity.” “We do not think there is fabrication,” he says.

Iwamoto now works at another hospital, Saya says. “He has a very nice reputation.” When I mention that I would like to talk to him, he suggests I should not. “He is very exhausted,” he says. “Better not to contact him at this moment.” Otherwise “the same thing” might happen that happened with Sato.

“What happened with Sato?” I ask. “People say he committed suicide over this,” Saya says. But he doesn’t know whether that’s true.

## V: ENIGMA

Sato’s fraudulent work has propelled him to No. 6 on Retraction Watch’s list of researchers who have racked up the most retractions. At the top is Japanese anesthesiologist Yoshitaka Fujii, with 183 retractions; his frequent co-author Yuhji Saitoh, also from Japan, is at 10th place, while Japanese endocrinologist Shigeaki Kato is No. 8. Iwamoto is at No. 9. That means half of the top 10 are Japanese researchers. Yet only about 5% of published research comes from Japan. What explains the number of prolific Japanese fraudsters?

Michie Sakamoto, who is leading another investigation at Keio University, into Iwamoto’s studies in animals, says it has to do with respect. “In Japan, we don’t usually doubt a professor,” he says. “We basically believe people. We think we don’t need strict rules to watch them carefully.” As a result, researchers faking their results may be exposed only

after they have racked up many publications.

Outside researchers may also be less likely to question anomalous results from Japan. Several early critics of Sato’s work say they thought at first that his unusual results might be due to something uniquely Japanese. One case in point: In 2003, Sato published a study on data from 40 patients with a very rare affliction named neuroleptic malignant syndrome, collected over 3 years. In a letter to the journal, a U.K. neurologist said he and his colleagues “could only recall two such cases in living memory”—but instead of casting doubt on the study, they said it was interesting that the syndrome seemed so prevalent in Japan.

But none of that explains why Sato decided to embark on his fraud—and nobody seems to be able to shed much light on that question. “Given the number of papers he published, he must have spent a very large amount of time on them,” Bolland says. “I don’t understand what his gain was. ...

***“In Japan, we don’t usually  
doubt a professor.  
We basically believe people.  
We think we don’t  
need strict rules to watch  
them carefully.”***

**Michie Sakamoto, Keio University**

There must have been some reason to do it.” The Keio University panel is just as puzzled. “We discussed this a lot in the committee,” Saya says. It might have been like a hobby, he suggests. A thrill. Saya uses the word “*otaku*,” a Japanese term often applied to people who read manga obsessively.

I thought I might find more clarity at the place where Sato perpetrated the fraud.

Mitate Hospital is not known for its scientific excellence. Except for one 2006 paper on schizophrenia, its entire research output over the past 20 years was produced by Sato.

The hospital is a sprawling complex of beige buildings set against green hills. I walk up to the reception. It’s quiet, no patients in sight. The receptionist does not understand me and asks me to write down what I want to say. “I am a journalist,” I write. “I would like to talk to the director about Yoshihiro Sato.”

Her eyes widen as she reads the name and she calls a nurse who speaks English. The nurse calls the director. “He does not want to talk to you,” she says after she hangs up. We stand awkwardly next to each at the reception desk, both embarrassed. It is clear

that everybody wants me to leave.

As I walk back to the bus stop I look back at the hospital. It is an unlikely place for an unlikely story. What did people here think about the research superstar in their midst? What do they remember about the man?

The effects of Sato’s fraud are still rippling out: citations, retractions, investigations. But the place at the epicenter of the disaster reveals nothing. Mitate Hospital squats silently in the midday sun.

## VI: EPILOGUE

Hours before I leave Japan, I meet Iwamoto’s lawyer, Satoshi Ogawa. We sit in the lobby of my Tokyo hotel, our words echoing from the bare walls and the marble floor. Ogawa says Iwamoto agreed to our conversation because he wants me to understand his point of view. “From his perspective, he is a victim.” Avenell’s team, says Ogawa, is now giving Iwamoto’s papers a level of scrutiny that is unfair and is causing his client a great deal of distress.

Ogawa says Sato wrote a detailed account of his interactions with Iwamoto a year before he died. He shows me an English version of the document, signed by Sato and witnessed by Ogawa and a notary. “I strongly requested Mr. Iwamoto to include my name as an author on the articles for which Mr. Iwamoto was the lead author,” Sato wrote. “I also started including Mr. Iwamoto’s name in the articles for which I myself was the lead author.”

The letter does not mention fraud, however. “I couldn’t force him to confess,” Ogawa says. “I think he had a mental illness.” His emails were not logical, he says. “To tell the truth, I predicted that he would commit suicide.”

*Suicide.* Is he sure that’s what happened?

“I received the information from the lawyer of Mr. Sato,” Ogawa says. Sato also left a note, he says, and he paraphrases it: “I am very sorry for Mr. Iwamoto. I decided to commit suicide.”

When I call Avenell after my return from Japan and tell her what I have learned, there is stunned silence at first. “That’s what we were dreading,” she says. “That’s horrible, really horrible.” Exposing the misconduct was important, she says. “Could we have done it without Sato committing suicide? So that he felt less guilty? I just don’t know.”

Later she follows up with an email, still astonished at “how such a small piece of data analysis a long time ago can end up with someone dying.” As a clinician and a researcher, Avenell wrote, she knows her work can eventually make the difference between life and death. “But seldom is the connection between a clinician and another human being’s death so obvious.” ■

# Plant nutrient acquisition entices herbivore

Maize defense metabolites enable iron uptake and attract herbivore attention

By Daniel J. Kliebenstein

To survive in highly complex environments, plants universally rely on specialized, or secondary, metabolites to withstand abiotic challenges (for example, wax to limit transpiration) and biotic challenges (for example, glucosinolates to deter herbivores). These metabolites are lineage-specific, and functional studies usually consider them to have a singular function. However, the complexity of the environment is much larger than the number of secondary metabolites within a plant, indicating that individual specialized metabolites may need to have multiple roles. As the number of functions of a single metabolite increases, so does the number of proteins and processes affected, and there is no guarantee that all of these interactions are positive. On page 694 of this issue, Hu *et al.* (1) show that benzoxazinoids, a textbook example of specialized metabolites in maize, have a functional duality centered around iron acquisition. The benzoxazinoids aid the uptake of iron. However, this generates a penalty because the herbivore, western corn rootworm, senses benzoxazinoid-iron complexes as a cue to locate and consume maize plants. This presents an evolutionary quandary, whereby maize can acquire iron and be eaten or maize can starve itself of iron and avoid herbivory—most plants have metabolites that create a similar conundrum.

Benzoxazinoids are critical for how maize interacts with a wide array of environmental components. They attract beneficial microbes to the rhizosphere (the region of soil

An adult western corn rootworm, *Diabrotica virgifera virgifera*, surveys potential food sources.



in contact with plant roots) to help shape the maize microbiome (2). They are also critical for defense against specialist and generalist herbivores through unknown mechanisms. However, benzoxazinoids are ineffective in defending against the western corn rootworm (3, 4) because this herbivore has developed the ability to modify and store benzoxazinoids as its own anti-parasitic drugs (anthelmintics). This allows the western corn rootworm to deter pathogenic nematodes (5). This interaction is not benzoxazinoid-specific and is a hallmark of most plant specialized metabolites, with numerous studies showing that individual metabolites have beneficial and detrimental roles that can span multiple trophic levels (6). For example, glucosinolates that are specific to the *Capparales* order provide defense against pathogens, insects, and avian herbivores (7, 8). Simultaneously, aphids can accumulate glucosinolates to provide defense against predators, which results in a net cost to the plant by promoting aphid survival (9). Similar opposing interactions are observed with the cardenolides of *Asclepias* (milkweeds), and the glycoalkaloids of the *Solanaceae* (nightshades) order (6). Thus, even though each individual metabolite is lineage-specific, the complex system of beneficial and detrimental functions across trophic levels is shared across plant lineages. This complexity of benefits and costs spread across potential abiotic and biotic interactions highlights the difficulty of adequately describing or measuring the fitness consequence of any specific metabolite.

Developing approaches to assess the fitness consequences is essential to understand the evolutionary pressures associated with the natural occurrence as well as the synthetic biology application of these natural pesticides into new plants. The study of Hu *et al.* shows that the fitness benefit attributed to a single role of any specialized metabolite is inadequate. Instead, a full description of the cost and benefit of that compound in every interaction in which it plays a role is required. This would involve cataloging the frequency with which each of the different environmental interactions may occur. The conditional interaction of multiple positive and negative functions with a fluctuating environment generates a complex fitness landscape. It is possible that the environment is sufficiently complex that there may not be one solution, but instead, individuals within a species may

have different metabolic solutions to maximize species fitness (10, 11).

Additionally, the mechanisms by which specialized metabolites function are largely unknown. It could be that opposing functions are the result of the metabolite having an independent mechanism in each interaction. For example, the specialized metabolite apigenin can bind a diverse range of proteins and variably affect their functions, with each interaction having potentially contradictory impacts on fitness (12). Alternatively, perhaps fitness costs and benefits arise from a single related mechanism as its function disseminates across the interaction of the plant with the environment. Hu *et al.* show that the cost and benefit of benzoxazinoids explicitly link to a single iron-binding (siderophore) function, which simultaneously allows iron uptake and invites the attack of a major herbivore. The authors propose a linear pathway whereby the plant may have first utilized these compounds to acquire iron and the rootworm then evolved the ability to track the compound to locate food.

The findings of Hu *et al.* will lead to exciting future studies necessary to assess if the siderophore function accounts for the majority of benzoxazinoid functionality. Alternatively, do benzoxazinoids have other, unrelated biological mechanisms? If a compound has multiple independent mechanisms, it is unlikely that these arose in a linear fashion; instead, they may represent evolutionary

serendipitous interactions. Serendipity raises complex issues for synthetic biology efforts to transfer specialized metabolites from one plant to another because unexpected interactions should be anticipated even in plants naïve to the metabolite (13). Hu *et al.* have taken the first step to understanding how specialized metabolite functions may have evolved and concurrently what engineering them involves. ■

#### REFERENCES

1. L. Hu *et al.*, *Science* **361**, 694 (2018).
2. A. L. Neal *et al.*, *PLOS ONE* **7**, e35498 (2012).
3. L. N. Meihls *et al.*, *Plant Cell* **25**, 2341 (2013).
4. C. A. M. Robert *et al.*, *Ecol. Lett.* **15**, 55 (2012).
5. C. A. M. Robert *et al.*, *eLife* **6**, e29307 (2017).
6. A. A. Agrawal, *J. Chem. Ecol.* **37**, 230 (2011).
7. J. Fan *et al.*, *Science* **331**, 1185 (2011).
8. P. W. Lambdon *et al.*, *Ann. Appl. Biol.* **133**, 313 (1998).
9. R. Chaplin-Kramer *et al.*, *J. Appl. Ecol.* **48**, 880 (2011).
10. R. E. Kerwin *et al.*, *New Phytol.* **215**, 1249 (2017).
11. R. Kerwin *et al.*, *eLife* **4**, e05604 (2015).
12. D. Arango *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, E2153 (2013).
13. F. G. Malinovsky *et al.*, *Plant Physiol.* **164**, 1443 (2014).

10.1126/science.aau6017

#### DEVELOPMENT

## Choosing the right input in cell signaling

A decoding module discriminates ligands and signaling activation in the Wnt pathway

By Kibeom Kim and Lea Goentoro

Over the evolutionary history of multicellular animals, several ancient signaling pathways have been conserved and reused for a myriad of functions. One such pathway is the Wnt signaling pathway, the almost ubiquitous roles of which include vital processes such as organogenesis, stem cell maintenance, and regeneration (1). This diversity of functions of the Wnt pathway may explain the diversity of ligands and receptors that activate these processes. In humans, there are 19 Wnt ligands, 10 Frizzled (FZD) receptors, and two lipoprotein receptor-related protein (LRP) co-receptors, as well as other receptors, such as RAR-related orphan receptor and receptor-like tyrosine kinase (2). Despite this diversity, binding interactions between Wnt ligands and FZD receptors are highly conserved (1), so that multiple Wnt ligands can bind to a FZD, and each Wnt ligand can bind to multiple FZDs to elicit cellular responses (3). Given this promiscuity, it is an ongoing pursuit to understand how cells can selectively respond to a specific Wnt ligand. On page 663 of this issue, Eubelen *et al.* (4) pursue this question and resolve the mechanism of how WNT7 selectivity is achieved. Resolving the mechanism for ligand discrimination has implications for understanding information flow in cell signaling and reveals new therapeutic opportunities.

Several strategies have been proposed for how cells could detect a specific Wnt ligand. Some studies showed that different FZD receptors show intrinsically varying binding affinities for different Wnt ligands (5). Thus, selectivity could be achieved to some extent by expressing the

Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91101, USA.  
Email: goentoro@caltech.edu



A western corn rootworm forages for maize roots.

FZDs with high affinity for a specific Wnt ligand. Another strategy for selectivity is through regulating the spatiotemporal expression of Wnt ligands, which is observed during development and in adult tissues (7). However, it is also typical for various Wnt ligands to be expressed in the same tissue. Another way to ensure specificity is through receptor cofactors. Two cell-membrane proteins, reversion-inducing cysteine-rich protein with Kazal motifs (RECK) and G protein-coupled receptor 124 (GPR124), selectively enhance WNT7 signaling in vascular development in vertebrate brains (6, 7). Although RECK and GPR124 were known to physically interact with each other and with FZD4 and LRP5 (6, 7), it was not known how their involvement confers selective binding to WNT7.

Eubelen *et al.* generated a human embryonic kidney cell line in which all 10 FZD

genes were inactivated. Indeed, the authors demonstrated that the intracellular domain of GPR124 binds to Dishevelled, a cytosolic adaptor that upon Wnt ligand–receptor binding is recruited to FZD. Because Dishevelled facilitates Wnt pathway signalosome assembly through its ability to polymerize (8), the authors hypothesized that GPR124 could function to bring together RECK–WNT7 to FZD–LRP. Consistently, they showed that GPR124 colocalizes with FZD4 in a Dishevelled-dependent manner.

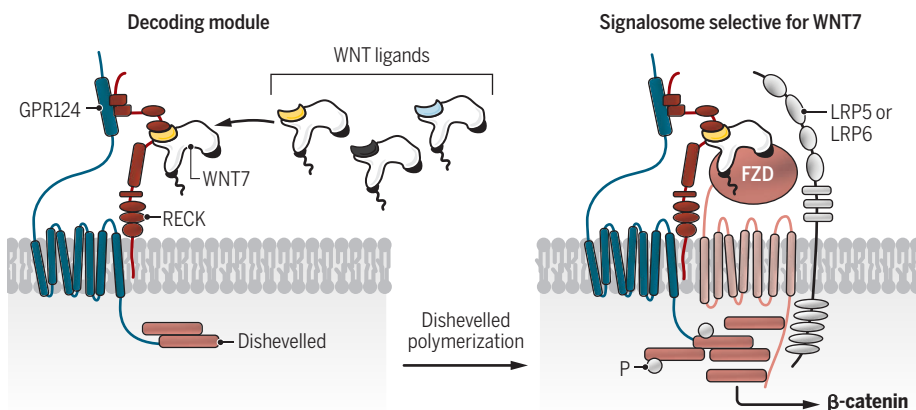
The elucidation of the mechanism by which RECK and GPR124 function as “a decoding module” of WNT7 by Eubelen *et al.* (see the figure) raises questions on how such ligand selectivity at the membrane is transduced into the nucleus. WNT7 signaling in brain angiogenesis is transduced downstream through activation of the  $\beta$ -catenin pathway (9). But many other

as well as the noncanonical pathways, such as the planar cell polarity and  $\text{Ca}^{2+}$  signaling (11), by engaging with different receptors (12). Moreover, Dishevelled plays a role in both the canonical and noncanonical pathways and interacts with multiple Wnt receptors (13, 14). Thus, considering these seemingly nonspecific uses of WNT7 and Dishevelled, it will be interesting to investigate how the mechanism uncovered for WNT7 ligand specificity is transduced into activation of select downstream pathways and, in particular, if the decoding module plays a role in this.

Evolving a modular mechanism for ligand specificity, separate from core signal transduction, may allow flexibility in fine-tuning the two processes independently for use in diverse contexts. Within this modular architecture, it is intriguing that a decoding module could have a dual function as a potentiator or inhibitor of signaling. Future investigations should examine whether the GPR124–RECK decoding module is used in other contexts, and whether there are different decoding modules for different Wnt ligands. Last, malfunctions in the Wnt pathway are found in diseases, including neurological disorders and cancer (15). Because Wnt signaling is critical in multiple organs, targeting core components of the pathway is likely to lead to multiple side effects (15). Despite ongoing clinical trials, no drugs targeting the Wnt pathway have been approved. However, having a decoding module that confers specificity to Wnt ligand signaling could lead to more selective targeted therapy. For example, cerebrovascular integrity is disrupted in various neurological diseases, including glioblastoma and stroke, and understanding the mechanism for specifically activating WNT7 signaling, which controls cerebrovascular development, holds exciting therapeutic promise. ■

## WNT7 signaling in brain vascular development

GPR124–RECK functions as a decoding module to specifically recruit WNT7 and present this ligand, in a Dishevelled-dependent manner, to the FZD–LRP5/6 receptor complex. WNT7 signalosome assembly leads to Dishevelled phosphorylation (P) and activation of  $\beta$ -catenin.



and two LRP genes were inactivated. In the absence of FZDs and LRPs, the authors found that RECK and GPR124 could not transduce WNT7 signaling, but that RECK physically binds to WNT7. By itself, RECK functions as a competitive inhibitor of signaling by sequestering WNT7. However, the presence of GPR124 converts RECK from inhibiting into potentiating WNT7 signaling. Therefore, RECK functions as a specificity factor that recruits WNT7 to the cell membrane, and GPR124 determines signaling activation.

How does GPR124 facilitate this WNT7-mediated signaling? Structure-function analysis in zebrafish brain angiogenesis revealed that the intracellular domain is necessary, but not the transmembrane domain. This suggests that GPR124 acts as a signaling-deficient receptor protein in WNT7 sig-

Wnt ligands—for example, WNT3A—can also signal through  $\beta$ -catenin. Why would a sophisticated mechanism for WNT7 specificity evolve, only for the signaling to converge downstream on  $\beta$ -catenin? It would be interesting to investigate how the selectivity achieved in Wnt ligand–receptor interaction is reflected in  $\beta$ -catenin activation. For instance, might the WNT7 signalosome produce  $\beta$ -catenin activation with different dynamics than that of the WNT3A signalosome? Such ligand-specific signaling dynamics has been demonstrated in the Notch pathway, in which delta-like ligand 1 (DLL1) produces pulsatile downstream signaling, whereas DLL4 produces sustained signaling, which leads to different transcriptional outcomes (10).

Many Wnt ligands, including WNT7, can activate the canonical  $\beta$ -catenin pathway

## REFERENCES

1. R. Nusse, H. Clevers, *Cell* **169**, 985 (2017).
2. C. Niehrs, *Nat. Rev. Mol. Cell Biol.* **13**, 767 (2012).
3. O. Voloshanenko *et al.*, *FASEB J.* **31**, 4832 (2017).
4. M. Eubelen *et al.*, *Science* **361**, eaat1178 (2018).
5. J. P. Dijksterhuis, J. Petersen, G. Schulte, *Br. J. Pharmacol.* **171**, 1195 (2014).
6. B. Vanhollebeke *et al.*, *eLife* **4**, e06489 (2015).
7. C. Cho, P. M. Smallwood, J. Nathans, *Neuron* **95**, 1056 (2017).
8. M. Gammons, M. Bienz, *Curr. Opin. Cell Biol.* **51**, 42 (2018).
9. J. M. Stenman *et al.*, *Science* **322**, 1247 (2008).
10. N. Nandagopal *et al.*, *Cell* **172**, 869 (2018).
11. F. Le Grand *et al.*, *Cell Stem Cell* **4**, 535 (2009).
12. K. M. Loh, R. van Amerongen, R. Nusse, *Dev. Cell* **38**, 643 (2016).
13. C. Gao, Y.-G. Chen, *Cell. Signal.* **22**, 717 (2010).
14. M. Nishita *et al.*, *Mol. Cell. Biol.* **30**, 3610 (2010).
15. M. Khan, *Nat. Rev. Drug Discov.* **13**, 513 (2014).

10.1126/science.aau6457



# Caged clusters shine brighter

The structure of small silver clusters confined in zeolites helps to explain their luminescent properties

By **Marta Quintanilla**<sup>1</sup> and  
**Luis M. Liz-Marzán**<sup>1,2</sup>

**M**etal clusters consisting of only a few atoms have different physical properties than single atoms and larger nanoparticles. Like molecules, they have discrete energy levels that allow electronic transitions; these transitions are the source of their distinct properties (1). To maintain their special features, clusters must be protected against aggregation and chemical modifications. Scientists have developed water-dispersible clusters that are stabilized with the help of short peptides, proteins, DNA, or polymers (2). Clusters may also be stabilized in solid-state matrices such as glasses, zeolites, or ultracold noble gases. On page 686 of this issue, Grandjean *et al.* (3) elucidate the origin of silver-cluster luminescence in a sodalite zeolite, a simple model that can withstand high-energy irradiation and displays bright green luminescence. They link the properties of the emitted light to different cluster structures in the zeolite.

The use of zeolites as scaffolds to encage silver clusters is particularly relevant for practical applications, because they can be produced reliably in industrial quantities at low cost. Zeolites have well-defined molecular-scale channels and cavities that are perfectly suited to host small clusters. Furthermore, the zeolite framework contains cation-exchange sites that facilitate incorporation of metal ions, which form clusters upon heat treatment or x-ray irradiation.

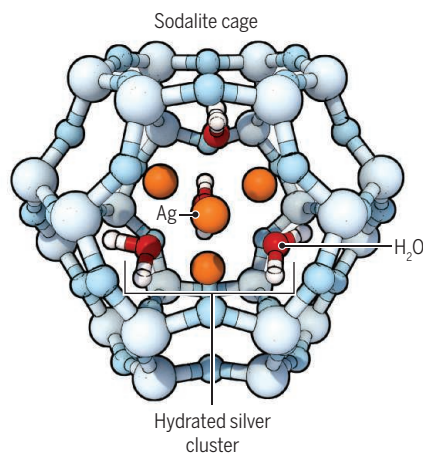
The luminescence of silver clusters stabilized in zeolites depends on the topology of the zeolite, thermal treatment, presence of charge-balancing ions ( $K^+$ ,  $Na^+$ , or  $Ca^{2+}$ ), and silver loading. Variation of these parameters gives access to luminescent emission under ultraviolet excitation in a broad range of wavelengths, from blue to near-infrared, and an external quantum yield (the conversion efficiency of incident light to emitted photons) that, in some cases, may be as high as 97% (4).

The high photostability and quantum yield of zeolite-encaged silver clusters make them

outstanding candidates for light-based applications. For example, fluorescent lamps, light-emitting diodes (LEDs), and organic LEDs use wavelength converters to control the color of the output light. In the case of commercial LEDs, there is still a need for the development of efficient converters that produce warm white light. This is a field in which zeolite-confined silver clusters appear as a highly competitive candidate (5). These materials have also been proposed as optical humidity sensors and luminescent tags (6).

## Confined clusters

Schematic representation of a hydrated  $Ag_4$  cluster inside a sodalite zeolite cage. The specific structure of this nanocomposite material is responsible for a bright photoluminescence, which can be tuned to produce a wide variety of colors.



Two main limitations restrict the development of applications for this exciting material. First, typical zeolite particle sizes (micrometers to tens of micrometers) compromise effective use as optical sensors and tags, which require smaller particles that can access remote areas of interest, for example, in biological imaging. Recently developed protocols yield zeolite particles with nanometer dimensions and high colloidal stability (7), but these protocols have not yet been applied to the silver-clusters technology.

The second limitation is the practical challenge of characterizing the structural and electronic properties of the material. Intricate host-guest interactions introduce interrelated effects that make it difficult to

draw general conclusions. Further, the sensitivity of both clusters and zeolites toward highly energetic probing beams limits the use of powerful techniques such as x-ray diffraction, extended x-ray absorption fine structure (EXAFS), and transmission electron microscopy (8). Photoluminescence experiments, combined with conventional structural-characterization techniques such as infrared and Raman spectroscopies, can partially disentangle the relations between structural features and luminescent properties. However, the results are, in some cases, contradictory, preventing a complete understanding of the system that could enable prediction, and thus tailoring, of the properties of silver clusters in zeolites (6).

Grandjean *et al.* now report a systematic analysis of silver clusters in a sodalite zeolite with x-ray excited optical luminescence EXAFS and time-resolved spectroscopy. These spectroscopic methods make it possible to selectively determine the structure of the emitting silver species involved in photoluminescence. The authors unambiguously identify complex hydrated silver clusters  $[Ag_4(H_2O)_x]^{2+}$  ( $x = 2$  and  $4$ ) at the center of sodalite cages (see the figure). This is an important finding in a field in which structural models have often been incomplete and sometimes contradictory. The complementarity of the experimental techniques chosen by the authors was essential to achieve this goal.

Theoretical modeling indicates that the remarkable optical properties resulting from this structure originate from a confined superatom quantum system in which hybridized silver and water oxygen orbitals are highly delocalized over the cluster. Upon excitation, one electron of the s-type highest occupied molecular orbital is promoted to the p-type lowest unoccupied molecular orbitals and relaxes through enhanced system intercrossing into long-lived triplet states.

The authors' approach may be extended to provide detailed understanding of the luminescent properties of other materials containing few-atom clusters made of silver and, perhaps, other metals. Such studies may yield a whole class of materials with tailored properties, based on luminescent clusters, that will certainly find technological applications in the near future. ■

## REFERENCES

1. H. Häkkinen, *Adv. Phys.* **X1**, 467 (2016).
2. S. Choi, J. Yu, *APL Mater.* **5**, 053401 (2017).
3. D. Grandjean *et al.*, *Science* **361**, 686 (2018).
4. O. Fenwick *et al.*, *Nat. Mater.* **15**, 1017 (2016).
5. W. Baekelant *et al.*, *ACS Energy Lett.* **2**, 2491 (2017).
6. E. Coutinho-Gonzalez *et al.*, *Acc. Chem. Res.* **50**, 2353 (2017).
7. H. Awala *et al.*, *Nat. Mater.* **14**, 447 (2015).
8. T. Altantzis *et al.*, *ACS Nano* **10**, 7604 (2016).

<sup>1</sup>CIC biomaGUNE and Centro de Investigación Biomédica en Red—Bioingeniería, Biomateriales, y Nanomedicina (CIBER-BBN), Paseo de Miramón 182, 20014 Donostia—San Sebastián, Spain. <sup>2</sup>Ikerbasque, Basque Foundation for Science, 48013 Bilbao, Spain. Email: llizmarzan@cicbiomagune.es

## IMMUNOLOGY

# Phase separation focuses DNA sensing

Liquid droplets form upon recognition of DNA to stimulate innate immune responses

By **Andrea Ablasser**

**T**he presence of any double-stranded DNA (dsDNA) in the cytosol is a potent trigger for the activation of innate immune responses. In the context of an infection, immunorecognition of “foreign” dsDNA serves as a central host strategy to detect the presence of pathogens. Under certain conditions, however, sensing of aberrant cytosolic self-dsDNA can also promote destructive inflammation in autoinflammatory diseases. A fundamental question is how cells spatially organize and biochemically regulate the recognition of intracellular dsDNA in the vast cytosolic space of a cell. On page 704 in this issue, Du and Chen (1) introduce an elegant explanation by demonstrating that phase separation of the dsDNA sensor cGAS [cyclic GMP (guanosine monophosphate)-AMP (adenosine monophosphate) synthase] enables this central task. This adds to the cellular processes that are facilitated by phase separation.

When it comes to innate immune detection, viruses are exquisitely challenging pathogens. Exploiting several host-cell processes for their own propagation, viruses usually express few pathogen-specific molecular signatures that could serve as motifs for pattern recognition receptors of the innate immune system. A central recognition strategy that virus-infected cells instead rely on is the detection of “mislocalized” nucleic acids. The nucleotidyltransferase cGAS, identified by the same laboratory in 2012 (2), is a pivotal innate immune receptor that detects pathogenic dsDNA. In response to binding dsDNA, cGAS catalyzes the synthesis of the second messenger molecule cGAMP (cyclic GMP-AMP), which triggers a series of cellular events orchestrated mainly by activating the endoplasmic reticulum-localized protein STING (stimulator of interferon genes) (3–5). A hallmark outcome of cGAS-cGAMP-STING signaling is the expression of various cytokines and

type I interferons that collectively elicit a powerful innate immune response.

Structural studies have provided important insights into the nature of the cGAS-DNA interaction at the molecular level, indicating the formation of a tetrameric (2:2) complex, which, on long dsDNA molecules, can further create ladderlike cGAS networks (6–8). Intriguingly, at the cellular level, activation of cGAS by dsDNA occurs in a length-dependent manner and is accompanied by the formation of discrete cGAS-containing cytosolic foci. However, it remained unclear how the formation of these aggregates could be explained in molecular terms and what functional properties they may

and thereby potentiate or dampen specific biological reactions. Mechanistically, phase separation is driven by multivalent interactions and, of note, can be facilitated by polymeric scaffolds such as nucleic acids, to which these proteins bind (13).

Du and Chen found that in the presence of dsDNA, cGAS demixes in solution and forms liquid-like droplets, which can fuse with one another to generate larger spherical units in vitro. The cGAS within these droplets is highly mobile, and, over time, the condensates acquire solid properties reminiscent of colloidal gels. Notably, in living cells, DNA-induced cGAS foci bear characteristics of phase-separated compartments in that they display fusion events with each other or allow rapid exchange of molecules into the droplets, including the cGAS substrates adenosine triphosphate (ATP) and guanosine triphosphate (GTP).

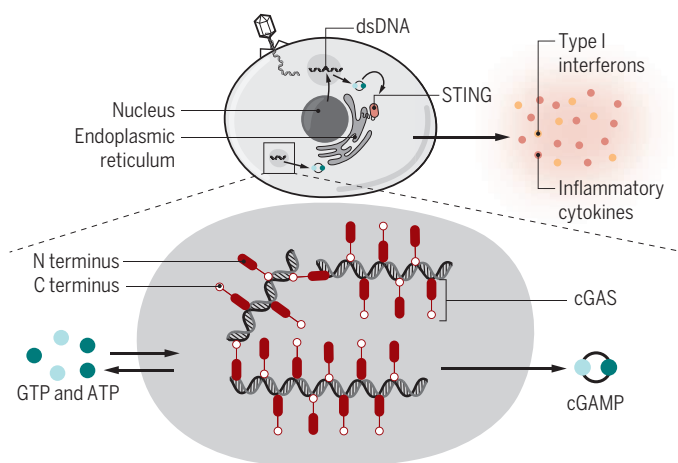
Could phase separation of cGAS regulate its enzymatic activation? The authors show that cGAMP synthesis correlates with the degree of phase separation in vitro and that, within cells, active cGAS is contained in droplets. These findings point to a model in which dsDNA nucleates cGAS, which is further concentrated into liquid-like droplets to promote the synthesis of cGAMP, which in turn activates STING at the endoplasmic reticulum (see the figure). Interestingly, Du and Chen

note that RNA and single-stranded DNA (ssDNA) also demix cGAS in vitro yet fail to activate cGAS. Thus, ligand specificity is not required to drive phase separation, but it defines the enzymatic activation of cGAS. Whether sequestering of cGAS via RNA- or ssDNA-induced demixing, as might occur during virus infection or erroneous accumulation of self-nucleic acids, negatively affects innate DNA sensing and the attendant immune responses remains an interesting issue that needs to be addressed.

Exploring the conditions under which the formation of cGAS droplets occur, Du and Chen discover the existence of sharp boundaries that demark a switchlike behavior of

## Phase separation promotes innate DNA signaling

Cytosolic dsDNA activates a potent innate immune response through cGAS activation. dsDNA nucleates cGAS, which is further concentrated into liquid-like droplets to promote the synthesis of cGAMP, which activates STING.



possess. Du and Chen now define them as membraneless compartments that are generated through a process of DNA-mediated liquid phase separation of cGAS.

Liquid-liquid phase separation is a physical process whereby components in a solution separate into two coexisting phases—one dense phase and one dilute phase. Biologically regulated phase separation has emerged as a key organizing principle underlying the formation of several membraneless compartments in living cells, including, for example, stress granules, the nucleolus, and P bodies in the cytoplasm (9–12). These liquid-like structures can concentrate or sequester specific biomolecules



phase separation. In addition to cGAS protein or dsDNA concentration, the length of the dsDNA molecules determines the threshold of cGAS unmixing. This observed length dependency reflects cellular responses to dsDNA (7). Interestingly, they find that, by providing multiple sites for intermolecular interaction with dsDNA, the disordered, positively charged N-terminal domain of cGAS enforces both the formation of lipid droplets as well as the production of cGAMP, emphasizing the importance of multivalence in promoting phase separation.

Phase separation is highly sensitive to changes in the chemical composition of the (cellular) environment. Du and Chen found that zinc ions, by stabilizing the DNA-cGAS complex, potentiate both droplet formation of cGAS as well as cGAMP production. More generally, these findings suggest that the cellular physicochemical environment may serve as a new checkpoint to regulate the DNA-sensing capability of cGAS and potentially other nucleic acid sensors. Understanding how the chemical composition of an environment modulates the phase-separating behavior of cGAS will be an interesting area for future investigation.

The study of Du and Chen offers a fresh perspective on the organization of innate immune signaling in general. What applies to cGAS may also be true for other cytosolic or membrane-bound receptors or adaptor molecules. Many intriguing questions remain for the future. What is the fate of cGAS droplets in living cells, and what are the molecular determinants that control this process? The study of Du and Chen used synthetic dsDNA. Thus, it will be interesting to assess the liquid phase-based organization of cGAS in different physiological scenarios, such as during pathogen infection or under conditions of endogenous DNA stress. Perhaps different pathogenic or endogenous DNA ligands or distinct biological contexts lead to differences in cGAS droplets with distinct cellular outcomes. Undoubtedly, it is important to understand the underlying principles governing the dynamic and layered discrimination of self- versus nonself-dsDNA by cGAS. ■

#### REFERENCES

1. M. Du, Z. J. Chen, *Science* **361**, 704 (2018).
2. L. Sun *et al.*, *Science* **339**, 786 (2013).
3. A. Ablasser *et al.*, *Nature* **498**, 380 (2013).
4. J. Wu *et al.*, *Science* **339**, 826 (2013).
5. G. N. Barber, *Nat. Rev. Immunol.* **15**, 760 (2015).
6. X. Li *et al.*, *Immunity* **39**, 1019 (2013).
7. X. Zhang *et al.*, *Cell Rep.* **6**, 421 (2014).
8. L. Andreeva *et al.*, *Nature* **549**, 394 (2017).
9. C. P. Brangwynne *et al.*, *Science* **324**, 1729 (2009).
10. C. P. Brangwynne *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 4334 (2011).
11. S. F. Banani *et al.*, *Nat. Rev. Mol. Cell Biol.* **18**, 285 (2017).
12. Y. Shin, C. P. Brangwynne, *Science* **357**, eaaf4382 (2017).
13. P. Li *et al.*, *Nature* **483**, 336 (2012).

10.1126/science.aau6019

#### ORGANIC CHEMISTRY

# Photocatalytic upgrading of natural gas

Cerium salts catalyze formation of value-added molecules from methane and small alkanes

By Motomu Kanai

**T**he recent shale gas revolution (1) continues to increase the estimates of the amount of natural gas both held in reserve and currently being produced. Natural gas contains methane (CH<sub>4</sub>) as the major component, in addition to ethane, propane, and butane as minor components. Natural gas is normally used as a fuel but could also be a cheap starting material for industrial chemical synthesis of value-added organic molecules (2), if ways are found to activate natural gas so that functional groups could be added. On page 668 of this issue, Hu *et al.* (3) report an important step toward such chemical processes. They upgraded natural gas to functionalized organic molecules by using a catalytic cerium salt and photoirradiation at ambient temperature.

Functionalization of alkanes must overcome the hurdles of low reactivity and poor selectivity (4). Cleavage of a C–H bond in alkanes for functionalization requires 105 kcal/mol. This high bond-dissociation energy (BDE) is one reason why natural gas is preserved under extreme conditions in Earth's crust on geological time scales. Chemical activation of alkanes through C–H bond cleavage generally requires harsh conditions, such as high temperature, superacid media, strong oxidants, or ultraviolet irradiation. Because the products and solvents are usually more reactive than alkanes, selective functionalization of the starting alkanes without product and solvent side reactions is difficult. Further, taking advantage of the scale of the natural-gas resource requires an inexpensive catalyst that is reactivated for many cycles, that is, one that exhibits very high catalyst turnover.

To overcome these difficulties, Hu *et al.* devised an alkoxy radical (RO•, where R = CH<sub>3</sub>, CCl<sub>3</sub>CH<sub>2</sub>, or CF<sub>3</sub>CH<sub>2</sub>) generated from a mixture of a cerium salt and an alcohol under blue light-emitting diode (LED) photoirradiation in an acetonitrile (CH<sub>3</sub>CN) solvent, as a catalyst for C–H bond cleav-

age of CH<sub>4</sub> and small alkanes (see the figure). Because an O–H bond is as strong as a C–H bond, the hydrogen-atom transfer process, which cleaves a CH<sub>4</sub> C–H bond by RO• and forms a methyl radical (•CH<sub>3</sub>) and an alcohol (RO–H), is thermodynamically neutral. This feature drives the reactivity of an alkoxy radical (RO•) in C–H bond cleavage of methane.

Selectivity is dictated by a polarity-matching effect (5). An alkoxy radical is electrophilic because of the high electron affinity of oxygen atoms. Thus, RO• prefers to generate a nucleophilic radical, such as a •CH<sub>3</sub>, because electrons smoothly flow from the C atom of CH<sub>4</sub> to the O atom of RO• in the transition state of hydrogen-

**“Compared with previous examples of catalytic functionalization of CH<sub>4</sub>... the method developed by Hu *et al.* proceeds under markedly mild conditions...”**

atom transfer from CH<sub>4</sub> (polarity match). In hydrogen-atom transfer from CH<sub>3</sub>CN, however, both RO• and •CH<sub>2</sub>CN are electrophilic. Thus, a competition between the two radicals for an electron on the abstracted H atom takes place that makes hydrogen-atom transfer from CH<sub>3</sub>CN by RO• unfavorable (that is, there is polarity mismatch), despite CH<sub>3</sub>CN containing an intrinsically weaker C–H bond (BDE = 93 kcal/mol) than in CH<sub>4</sub>. In this way, the radical reaction avoids the unproductive pathway of reaction with the solvent.

How can this reactive and selective alkoxy radical be generated? The answer by Hu *et al.* relies on photoactivation (6). A variety of reaction patterns have been realized with hybrid catalyst systems comprising a photosensitizer that absorbs light energy and promotes redox reactions between catalysts and reagents (7, 8). Hu *et al.* used the photoinduced ligand-to-metal

Graduate School of Pharmaceutical Sciences, The University of Tokyo, Tokyo 113-0033, Japan. Email kanai@mol.f.u-tokyo.ac.jp

charge transfer of a tetravalent cerium alkoxide ( $\text{ROCe}^{\text{IV}}$ ) to generate an alkoxy radical (9). Irradiation with blue light [wavelength ( $\lambda$ ) = 400 nm] excites the cerium alkoxide catalyst, and homolytic cleavage of the  $\text{ROCe}^{\text{IV}}$  bond generates the key alkoxy radical ( $\text{RO}^\bullet$ ) and a one-electron reduced trivalent  $\text{Ce}^{\text{III}}$  salt.

The  $\text{RO}^\bullet$  selectively cleaves a C–H bond of  $\text{CH}_4$  through hydrogen-atom transfer, generating  $\cdot\text{CH}_3$ . Alkyl radicals, including  $\cdot\text{CH}_3$ , are highly reactive and react with a N=N, C=C, or C=N bond of the other substrate; di-*tert*-butyl azodicarboxylate is shown as a representative substrate in the figure. The resulting radical intermediate is reduced by  $\text{Ce}^{\text{III}}$ , regenerating the  $\text{Ce}^{\text{IV}}$  alkoxide catalyst and affording functionalized products of  $\text{CH}_4$ . In a representative example,  $\text{CH}_4$  reacted with di-*tert*-butyl azodicarboxylate and was converted to an *N*-methylhydrazine derivative with a catalyst turnover number of 2900, meaning that one molecule of a simple cerium alkoxide catalyst can generate 2900 value-added organic molecules starting from  $\text{CH}_4$ . The reaction can also be performed in a continuous-flow system, enabling easy handling of gas-phase alkane substrates, enhanced efficiency in light utilization, and straightforward scale-up (10).

Compared with previous examples of catalytic functionalization of  $\text{CH}_4$  (11) and other alkanes, the method developed by Hu

*et al.* proceeds under markedly mild conditions (room temperature and near neutral pH). Thus, this method may be suitable for application to the synthesis of a variety of complex molecules, including drugs. Both functional group tolerance and control of stereoselectivity are essential for development in this direction, but these points remain unaddressed at this point. Additionally, the catalyst turnover efficiency and quantum yield (currently, ~0.2%) must be improved, considering the large scale needed for the processing of natural gas. Despite these potential hurdles, the achievement by Hu *et al.* is a step forward in using carbon feedstock natural gas as a starting material in the creation of organic molecules that benefit our society. ■

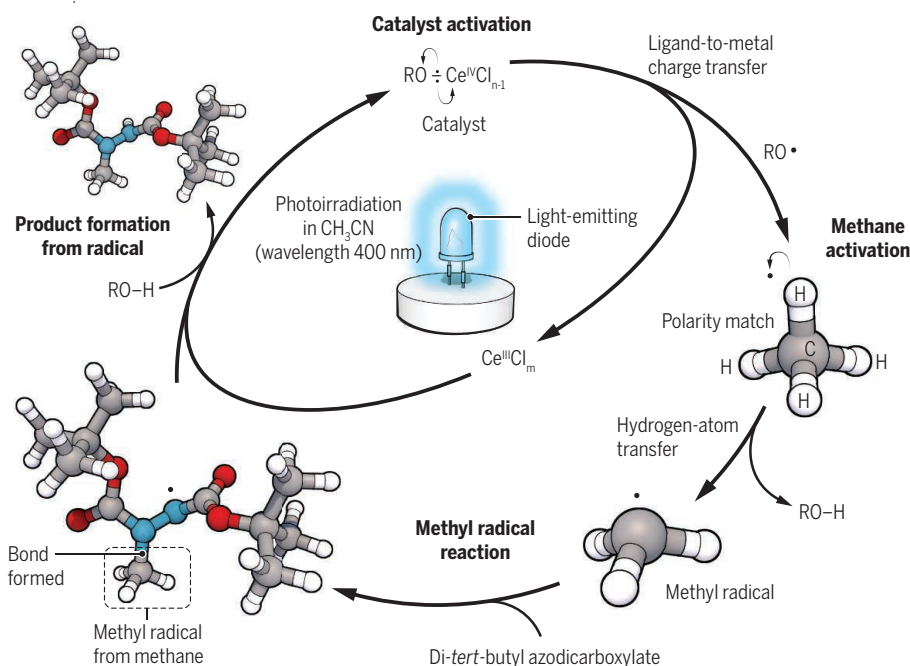
## REFERENCES

1. U.S. Energy Information Administration, International Energy Outlook 2017; [www.eia.gov/outlooks/ieo](http://www.eia.gov/outlooks/ieo).
2. K. C. Nicolaou, T. Montagnon, *Molecules That Changed the World* (Wiley-VCH, 2008).
3. A. Hu *et al.*, *Science* **361**, 668 (2018).
4. A. E. Shilov, G. B. Shul'pin, *Chem. Rev.* **97**, 2879 (1997).
5. B. P. Roberts, *Chem. Soc. Rev.* **28**, 25 (1999).
6. C. R. J. Stephenson, T. P. Yoon, D. W. C. MacMillan, *Visible Light Photocatalysis in Organic Chemistry* (Wiley, 2018).
7. M. H. Shaw *et al.*, *J. Org. Chem.* **81**, 6898 (2016).
8. A. Studer, D. P. Curran, *Angew. Chem. Int. Ed.* **55**, 58 (2016).
9. A. Hu *et al.*, *J. Am. Chem. Soc.* **140**, 1612 (2018).
10. D. Cambié, C. Bottecchia, N. J. Straathof, V. Hessel, T. Noel, *Chem. Rev.* **116**, 10276 (2016).
11. N. J. Gunsalus *et al.*, *Chem. Rev.* **117**, 8521 (2017).

10.1126/science.aau5379

## Photocatalytic upgrade of methane

The major component of natural gas, methane, is abundant but chemically inert. A photocatalytic method (shown) can activate and functionalize methane as well as other small alkanes, under mild conditions.



## DEVELOPMENT

## Turning up translation in fragile X syndrome

Fragile X mental retardation protein promotes translation, contrary to its assumed function

By Sameer Aryal<sup>1,2</sup> and Eric Klann<sup>2,3</sup>

**F**ragile X syndrome (FXS) is the most prevalent inherited form of intellectual disability and autism (1). FXS is usually caused by transcriptional silencing of the fragile X mental retardation 1 (*FMR1*) gene, which encodes fragile X mental retardation protein (FMRP), an RNA-binding protein that is thought to repress the translation of specific messenger RNAs (mRNAs). Precise translational control is especially critical in neurons because rapid synthesis of proteins from preexisting mRNAs underlies many forms of synaptic plasticity, which is altered in animal models of FXS (2). Progress has been made in identifying RNAs that FMRP binds, but its functional targets and modes of translational control remain elusive, especially during development (3–7). On page 709 of this issue, Greenblatt and Spradling (8) use *Drosophila melanogaster* oocytes to demonstrate that FMRP increases the translation of multiple long mRNAs, many of which are implicated in intellectual disability and autism.

Several lines of evidence have suggested that FMRP inhibits the translation of certain mRNAs. Increased net de novo protein synthesis is observed in multiple brain regions of mouse models of FXS and in cells derived from FXS patients (1). Moreover, the expression of several proteins encoded by FMRP-binding mRNAs are overexpressed in the brains of mice that lack FMRP (2). However, some evidence also indicates that the loss of FMRP is as-

<sup>1</sup>Sackler Institute of Graduate Biomedical Sciences, New York University (NYU) Langone Medical Center, New York, NY 10016, USA. <sup>2</sup>Center for Neural Science, NYU, New York, NY 10003, USA. <sup>3</sup>NYU Neuroscience Institute, NYU Langone Medical Center, New York, NY 10016, USA. Email: [eklann@cns.nyu.edu](mailto:eklann@cns.nyu.edu); [sameer.aryal@med.nyu.edu](mailto:sameer.aryal@med.nyu.edu)



sociated with decreased synthesis and/or expression of certain proteins encoded by FMRP-binding mRNAs (2, 5, 9), suggesting that it can also function to enhance mRNA translation.

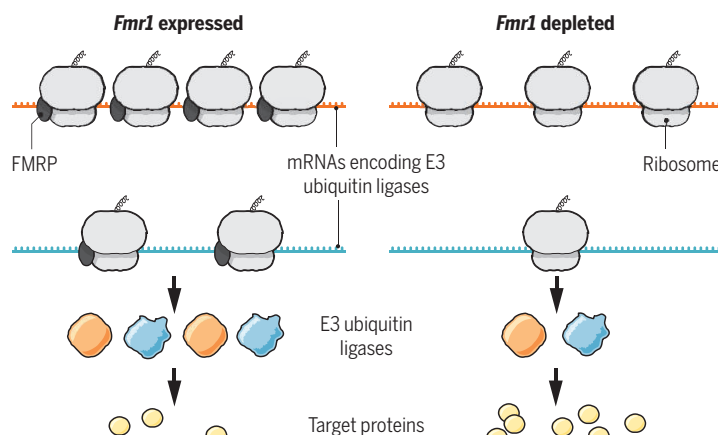
Building on investigations to identify mRNAs that bind to FMRP (3), genome-wide studies to identify translating mRNAs from hippocampal CA1 pyramidal neurons from FXS mice indicate that the translation of a substantial fraction of mistranslated mRNAs is down-regulated (6, 7). Surprisingly, there is limited overlap between the mistranslated mRNAs in neurons from FXS mice and those identified as FMRP-binding mRNAs. There are several possible explanations, which include the following: FMRP regulates another aspect(s) of metabolism of its mRNA targets; the mistranslating mRNAs are compensatory responses to the lack of FMRP (6); the loss of repression of FMRP target mRNAs causes secondary effects on translational control pathways, which then alter translation (1); and FMRP regulates translation in a cell type- and/or development-specific manner.

Because mature *D. melanogaster* oocytes lack transcription, they must maintain the translation of preexisting mRNAs at a distance from the nucleus, similar to neurons. Additionally, female flies can be manipulated to store mature oocytes rather than ovulate (10). This permits the investigation of functional FMRP targets in cells that rely exclusively on translational control for extended periods of time. The authors found that germ cell-specific depletion of *Fmr1* mRNA impaired the developmental capacity of stored oocytes in a time-dependent manner: *Fmr1*-deficient oocytes stored for a single day hatched normally. However, only 20 to 30% of oocytes hatched after 10 days of storage, and the resulting embryos showed defects in neural development. This suggests that FMRP plays a crucial role in maintaining the developmental competence of oocytes when they are dependent on the translation of preexisting mRNAs.

Greenblatt and Spradling carried out ribosome profiling on oocytes stored for 1 to 2 days to pinpoint the positions of translating ribosomes on individual mRNAs (11). They identified more than 400 genes that were differentially translated when *Fmr1* was depleted. Surprisingly, more than 95% of the mistranslated genes showed

## FMRP enhances translation

FMRP maintains the developmental capacity of *D. melanogaster* oocytes by promoting the translation of specific, mostly long, transcripts. The functional targets of FMRP, identified by reduction in ribosome loading with *Fmr1* depletion, are enriched in genes encoding E3 ubiquitin ligases, including POE.



decreased ribosome occupancy without alterations in mRNA abundance, suggesting that FMRP functions to enhance, rather than repress, the translation of many genes in *D. melanogaster* oocytes. More than 10% of the mistranslated genes were orthologs of human genes that have been implicated in neuropsychiatric disorders, including intellectual disability and autism. This finding resonates with the long-held belief that FMRP regulates the translation of multiple autism-linked genes (3). Deeper examination of the functions of these genes may aid in understanding the molecular basis for the comorbidity and variation in severity of FXS and autism.

The authors noted another remarkable feature of the mistranslating transcripts: They were longer than average. Nearly half of all mRNAs coding for long proteins (>1800 amino acids) had reduced ribosome loading when *Fmr1* was depleted. Notably, dysregulation of long genes has been observed in multiple neuropsychiatric disorders associated with synaptic dysfunction, including FXS (12) and autism (13). Thus, it is critical to examine why there is a convergence on the inability to appropriately express long genes in such disorders.

How are these findings reconciled with the previously assumed function of FMRP as a repressor of translation? Because FMRP expression in both mouse and *D. melanogaster* peaks during early development, it is possible that the enhancement of translation by FMRP is development specific. The authors speculate that the decreased synthesis of several E3 ubiquitin ligases, which target proteins for degradation, after FMRP reduction increases the

life span of translation factors that mediate increased overall de novo protein synthesis (see the figure). Indeed, the authors demonstrate that oocytes depleted of the E3 ubiquitin ligase protein purity of essence (POE), a translational target of FMRP, mimic those lacking *Fmr1*. This indicates that E3 ubiquitin ligases may be therapeutic targets for the treatment of FXS.

Given the overlap between mistranslating transcripts in oocytes and FMRP-binding mRNAs identified in mouse brains, it is possible that FMRP regulates the translation of some mRNAs by direct binding. However, further experiments are needed to better understand how FMRP establishes RNA-binding specificity.

Additionally, caution must be exercised when interpreting results that provide estimates of protein synthesis via ribosome profiling. Because FMRP is thought to impede ribosome translocation, its loss may cause ribosomes to transit at increased rates on specific mRNAs, resulting in decreased steady-state ribosome loading but increased de novo translation of those mRNAs. Although the authors rule this out for some mRNAs by examining their protein abundance, it is essential that whole-genome approaches be used to examine how FMRP modulates ribosome transit on individual mRNAs. Greenblatt and Spradling's findings add an intriguing new piece to the puzzle of identifying the functional targets of FMRP. In the future, it may be possible to identify a core set of mistranslated mRNAs that mediate pathophysiology in FXS and other forms of autism and intellectual disability. ■

## REFERENCES AND NOTES

1. J. C. Darnell et al., *Nat. Neurosci.* **16**, 1530 (2013).
2. M. R. Santoro et al., *Ann. Rev. Pathol. Mech. Dis.* **7**, 219 (2012).
3. J. C. Darnell et al., *Cell* **146**, 247 (2011).
4. M. Ascano et al., *Nature* **492**, 382 (2012).
5. R. Tabet et al., *Proc. Natl. Acad. Sci. U.S.A.* **113**, E3619 (2016).
6. S. R. Thomson et al., *Neuron* **95**, 550 (2017).
7. L. Ceolin et al., *Front. Mol. Neurosci.* **10**, 340 (2017).
8. E. J. Greenblatt, A. C. Spradling, *Science* **361**, 709 (2018).
9. C. Gross et al., *J. Neurosci.* **31**, 5693 (2011).
10. C. A. Jeffreys et al., *Curr. Biol.* **13**, 498 (2003).
11. N. T. Ingolia et al., *Science* **324**, 218 (2009).
12. R. L. Ouwenga, J. Dougherty, *Mol. Autism* **6**, 16 (2015).
13. M. J. Zylka et al., *Neuron* **86**, 353 (2015).

## ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grants NS034007, NS047384, and HD082013.

10.1126/science.aau6450





## POLICY FORUM

### CLIMATE CHANGE

# Global warming policy: Is population left out in the cold?

Population policies offer options to lessen climate risks

By John Bongaarts<sup>1</sup> and Brian C. O'Neill<sup>2,3</sup>

**W**ould slowing human population growth lessen future impacts of anthropogenic climate change? With an additional 4 billion people expected on the planet by 2100, the answer seems an obvious “yes.” Indeed, substantial scientific literature backs up this intuition. Many nongovernmental organizations undertake climate- and population-related activities, and national adaptation plans for most of the least-developed countries recognize population growth as an important component of vulnerability to climate impacts (1). But despite this evidence, much of the climate community, notably the Intergov-

ernmental Panel on Climate Change (IPCC), the primary source of scientific information for the international climate change policy process, is largely silent about the potential for population policy to reduce risks from global warming. Though the latest IPCC report (2) includes an assessment of technical aspects of ways in which population and climate change influence each other, the assessment does not extend to population policy as part of a wide range of potential adaptation and mitigation responses. We suggest that four misperceptions by many in the climate change community play a substantial role in neglect of this topic, and propose remedies for the IPCC as it prepares for the sixth cycle of its multiyear assessment process.

Population-related policies—such as offering voluntary family planning services as well as improved education for women and girls—can have many of the desirable

characteristics of climate response options: benefits to both mitigation and adaptation, co-benefits with human well-being and other environmental issues, synergies with Sustainable Development Goals (SDGs), and cost effectiveness. These policies can also enable women to achieve their desired family size, and lead to lower fertility and slower population growth (3). The resulting demographic changes can not only lessen the emissions that drive climate change but also improve the ability of populations to adapt to its consequences.

### MISPERCEPTION 1: POPULATION GROWTH IS NO LONGER A PROBLEM

The population growth rate of the developing world increased sharply in the 1950s and 1960s, resulting in a doubling of the world population from 3 billion in 1960 to over 6 billion in 2000 (4). The main cause of this acceleration was the spread of public health measures, which rapidly reduced death rates while birth rates remained high. Slowing this population expansion became a top priority for the global development agenda. In the 1970s and 1980s, substantial international assistance was invested in voluntary family planning programs to reduce fertility.

This early consensus on population policy ended in the 1990s when interest and international support waned for rea-

<sup>1</sup>Population Council, New York, NY, USA. <sup>2</sup>National Center for Atmospheric Research, Boulder, CO, USA. <sup>3</sup>University of Denver, Denver, CO, USA. Email: jbongaarts@popcouncil.org



A community health worker shows women how to use a condom in Bangladesh.

sons including (i) the belief that fertility declines already under way in Asia and Latin America would soon occur in Africa; (ii) the expectation that high AIDS mortality would halt population growth in sub-Saharan Africa; (iii) the failure of earlier dire predictions, such as worldwide famine, to materialize; and (iv) the call from the 1994 International Conference on Population and Development (ICPD) to emphasize reproductive health and rights over demographic aims. As a result, funding for reproductive health issues (e.g., maternal care, safe delivery, sexually transmitted diseases, and female genital cutting) rose and funding for family planning programs declined. These issues were widely debated in the 1990s and 2000s and may have affected the coverage of population policy in IPCC reports in part because population growth no longer seemed an urgent matter.

Over the past decade, two unexpected developments have led to renewed concern about future population growth, particularly in sub-Saharan Africa. Fortunately, AIDS mortality has dropped sharply as treatment has become more accessible worldwide. In addition, and contrary to expectations, birth rates across sub-Saharan Africa have remained high, and declines in birth rates have stalled in several countries. As a result, the latest UN world population projection is the highest ever, expecting 11.2 billion in 2100, a nearly 4 billion rise from 2015 (4). Much of this rise is projected in sub-Saharan Africa (from 1 billion in 2015 to 4 billion in 2100), but Asia (excluding East Asia) and Latin America are also projected to grow substantially.

## MISPERCEPTION 2: POPULATION POLICIES ARE NOT EFFECTIVE

Population policies generally recommend a range of interventions that influence fertility trends, either directly or indirectly. Family planning programs to assist women in achieving their reproductive goals for limiting or spacing births have been the main population intervention adopted by governments (5). These programs have been successful in a number of countries, but further investments are still needed. Each year, about 85 million unintended pregnancies result in 32 million unplanned births worldwide; the large majority of these (28 million) in the developing world (6). Population growth could be reduced substantially by avoiding these unplanned pregnancies.

The vast majority of unintended pregnancies occur among women who want to avoid pregnancy but are not using effective contraception. Reasons for non-use include

lack of access to services and the high costs of modern methods. Fear of side effects of methods, disapproval of partners, and reluctance to violate social norms are also substantial barriers to use (7).

Voluntary family planning programs designed to be responsive to cultural customs reduce these obstacles by increasing access, providing subsidies, and expanding method options. Well-run voluntary programs have contributed to sustained declines in fertility and population growth across Asia, the Middle East, and Latin America and in some countries in Africa (5, 7).

Although many studies have assessed outcomes of family planning programs, these mostly examined near-term effects on raising contraceptive use and reducing birth rates. Very few studies estimate impacts of family planning programs on longer-term demographic trends, in part because isolating the exact impact of programs versus socioeconomic development and other factors is not straightforward in most countries. The potential impact of these programs on long-range population size is illustrated by comparing Bangladesh and Pakistan, which had almost the same population size in 1980. Bangladesh then implemented one of the world's most effective voluntary family planning programs. By contrast, Pakistan's program was considerably weaker, lacking government commitment. As a result, fertility trajectories differed substantially from 1980 onward, resulting in increasingly large differences in population size over time (see the figure). By 2100, Pakistan's population is projected to be double the size of Bangladesh's. This suggests that the Bangladesh family planning program led to a large cut in the country's potential 2100 population. Fertility and population trends are also affected by levels of socioeconomic development, but this is unlikely to be the dominant explanation for the different trajectories (see supplementary materials). Development levels as measured by the Human Development Index are nearly the same for Bangladesh and Pakistan (8), which are both poor majority-Muslim countries in South Asia.

High-quality voluntary family planning programs can also have a large impact in Sub-Saharan Africa. Programs implemented in the early 2000s in Ethiopia, Malawi, and Rwanda have already resulted in sharp declines in fertility. Unfortunately, in many African countries, including Nigeria, family planning is still given low priority. Existing programs in Asia and Latin America also could be improved.

In addition to improving the health and welfare of women, families, and communities, reduced fertility assists in eliminat-

ing poverty and reducing pressure on the natural environment. Family planning programs are one of the most cost-effective health and development investments available to governments (9).

Population policies often include other relevant programmatic interventions such as education and empowerment of women. These are not only important development interventions, but they also accelerate fertility decline (10). Of course, educated and empowered women must have access to contraception to regulate their fertility.

## MISPERCEPTION 3: POPULATION DOES NOT MATTER MUCH FOR CLIMATE

The emissions and land use that drive climate change are a result of income and consumption growth, technological change, changes in economic structure, related policies, and other factors. Past and current emissions have been attributed primarily to economic growth powered by fossil fuels in the currently high-income countries. However, multiple studies using increasingly sophisticated methods have demonstrated that population plays an important role both in historical and projected future emissions (11). Although slower future population growth would not be the most important means of reducing future emissions, it could reduce global emissions by 40% or more in the long term (12). Slower population growth and associated changes in age structure can also have positive economic effects that would tend to drive greenhouse gas emissions up, but these effects do not appear to be large enough to offset the emissions reduction produced by the slower population growth (see supplementary materials).

The potential emissions reduction is large even though policy-induced declines in population growth would be largest in regions that currently have low per-capita emissions. Anticipated future growth in incomes and energy use in these developing regions explain this result. Over the next few decades, overall emissions from low-income countries are likely to rise because of a rise in emissions per capita from rapid industrialization, as well as because of increasing population.

Slower population growth is also anticipated to reduce climate change risks by freeing up resources for adaptation. Improvements in education and health, which can both lead to and result from slower population growth, can reduce vulnerability to natural disasters and climate risks (13). Population-related policies that affect the spatial distribution of population, including those influencing immigration, labor mobility, urbanization, and coastal development, can also affect vulnerability.

The IPCC itself has partially assessed the topic, concluding that population growth, urbanization, and changes in age structure are important drivers of emissions. It has also concluded that demography shapes the exposure and vulnerability of populations to climate impacts and can limit, or facilitate, the ability of society to adapt to those impacts. What is missing is an accounting of how reductions in population growth might play a role in responses to the climate issue (see supplementary materials).

Many in the climate policy community currently focus on achieving substantial emissions reductions in the near future. Although slowed population growth would contribute only modestly in the short term, its cumulative effect over the 21st century would be substantial. Slowed population growth would reduce emissions and the demand for energy that would have to be satisfied with low- or zero-carbon sources. It would therefore also have an important effect on the scale of the energy system ultimately required under a stabilized climate.

#### MISPERCEPTION 4: POPULATION POLICY IS TOO CONTROVERSIAL TO SUCCEED

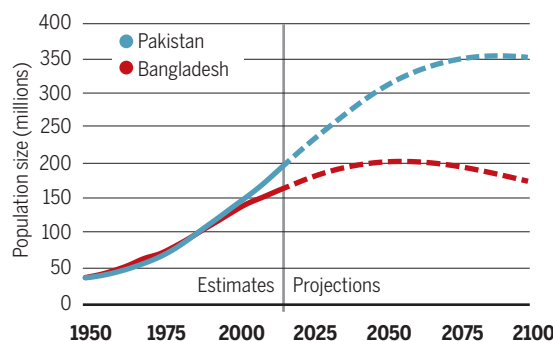
Family planning programs have attracted criticism since their initiation. The most persistent opposition has come from conservative religious and social groups. One of their main concerns is that making contraception readily available encourages promiscuity and leads to a breakdown of family life. In addition, the Catholic Church opposes artificial contraception, and Islam opposes sterilization. These religious concerns are shared by conservative political allies, leading to frequent controversy.

Other concerns about family planning programs have been raised by human rights advocates who fear coercion. They point to examples of massive abuses by the Chinese government during the implementation of the one-child policy and by the Indian government during an emergency period in the late 1970s. These abhorrent practices were and are universally condemned. Nevertheless, reproductive health remains a political issue in many countries, and constraints on women's choices continue to exist (e.g., limited choice of contraceptive methods or lack of services).

A key point of sensitivity is that family planning programs largely aim to reduce fertility in the developing world while people in the developed world, which is primarily responsible for causing the climate to change, continue their excessive emission of greenhouse gases. At the same time, many

## Population estimates and projections for Bangladesh and Pakistan

Differences suggest that a good family planning program (as in Bangladesh beginning in the 1980s) can have a large impact on population trajectories in the long term. Data are from (4). See supplementary materials for details on data and methods.



developed countries are also increasingly concerned about aging and decline of their populations. Many in the climate change community believe that entering into a population policy discussion thus blames the poor countries for problems created by the rich countries. Although this belief is real, it does not change the fact that population growth in developing countries poses challenges for climate and development and deprives the international community of an important policy lever to improve human welfare.

Although these controversies do indeed exist, they are not the obstacles to program implementation that some in the climate community believe them to be. Governments around the world now support the conclusions of the ICPD, confirmed by the SDGs, which call for a human rights-based approach and for women everywhere to have the right to freely choose when and how often to get pregnant (14). Many countries in Asia and Latin America have invested in family planning programs, and increasing numbers of governments in sub-Saharan Africa have started such programs. There is widespread agreement among governments and international organizations that family planning programs are a valuable investment. The SDGs in fact call for more such services. But these programs are often given low priority because they are considered a health investment rather than an investment with wide-ranging socioeconomic and environmental benefits.

#### POLICY LEVERS

Rapid population growth is one of the key drivers of emissions and one of the determinants of vulnerability to impacts; it therefore should be considered as a potential climate policy lever. A key first step in remedying the current neglect of the issue is for the IPCC to include population policy in its assessment

of the literature on mitigation and adaptation options. Although the outline for the sixth IPCC assessment report has already been agreed upon (with no explicit mention of population policy), there is ample opportunity within its structure to assess literature on population policy as a component of mitigation or adaptation responses, as well as its costs and benefits, implementation barriers, and links to SDGs (see supplementary materials). The IPCC should also consider the inclusion of more social scientists experienced in reproductive health and population policy.

Beyond the IPCC, the climate and environmental communities and international development institutions should embrace scientifically sound analyses of population policy and human rights-based reproductive health programs.

Other international environmental assessments (11, 15) have done somewhat better than the IPCC in covering this topic. Given the urgency of addressing climate change, all available options, especially those that have multiple benefits for sustainable development, should be assessed by experts and considered by governments. ■

#### REFERENCES

1. C. Mutunga, K. Hardee, *Afr. J. Reprod. Health* **14**, 133 (2010).
2. Intergovernmental Panel on Climate Change (IPCC), *Climate Change 2014: Synthesis Report* (IPCC, Geneva, Switzerland, 2014).
3. G. J. Abel et al., *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14294 (2016).
4. United Nations, *World Population Prospects: The 2017 Revision* (United Nations Population Division, New York, 2017).
5. J. May, *World Population Policies: Their Origin, Evolution and Impact* (Springer, 2012).
6. G. Sedgh, S. Singh, R. Hussain, *Stud. Fam. Plan.* **45**, 301 (2014).
7. J. Bongaarts, J. Cleland, J. Townsend, J. Bertrand, M. Das Gupta, *Family Planning Programs for the 21st Century: Rationale and Design* (Population Council, New York, 2012).
8. United Nations Development Program (UNDP), *Human Development Report 2016: Human Development for Everyone* (UNDP, New York, 2016).
9. Copenhagen Consensus Center, "The Economist: Special Online Supplement" (Copenhagen Consensus Center, Copenhagen, 2015); [www.copenhagenconsensus.com/post-2015-consensus/economist](http://www.copenhagenconsensus.com/post-2015-consensus/economist).
10. S. Jejeebhoy, *Women's Education, Autonomy, and Reproductive Behaviour: Experience from Developing Countries* (Clarendon, Oxford, 1995).
11. Millennium Ecosystem Assessment (MA), *Ecosystems and Human Well-Being: Synthesis* (Island Press, Washington, DC, 2005).
12. B. O'Neill et al., *Lancet* **380**, 157 (2012).
13. W. Lutz, R. Muttarak, E. Striessnig, *Science* **346**, 1061 (2014).
14. United Nations, Report of the International Conference on Population and Development (Population Division, Department for Economic and Social Information and Policy Analysis, A/CONF.171/13/Rev.1, United Nations, New York, 1995).
15. United Nations Environment Programme (UNEP), *Global Environment Outlook 5: Environment for the Future We Want* (United Nations Environment Programme, Nairobi, Kenya, 2012).

#### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/361/6403/650/suppl/DC1](http://www.sciencemag.org/content/361/6403/650/suppl/DC1)

10.1126/science.aat8680





The waters on which many residents depend for crab fishing may soon render Tangier Island uninhabitable.

BOOKS *et al.*

## CLIMATE POLICY

# An island on the brink

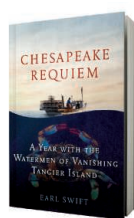
Convinced erosion, not climate change, threatens their home, a community grapples with an uncertain future

By Andrew J. Miller

Tangier Island, located in the middle of Chesapeake Bay, is one of the mainstays of the region's blue crab fishery. No location on the small island is more than 5 feet above sea level, and two-thirds of its land mass has been lost since 1850. Much of its upland is being converted to tidal marsh as sea-level rise accelerates while the land subsides at one of the highest rates in the mid-Atlantic region. In 2015, Schulte *et al.* argued that without government action, the citizens of Tangier Island "may become among the first climate change refugees in the continental USA" (1). It is anticipated that the island may become uninhabitable within 50 years and, given more recent updates in anticipated rates of sea-level rise, perhaps within 25 years.

Earl Swift picks up this theme in *Chesapeake Requiem*, warning that the very waters on which Tangier Islanders depend represent a threat to the community's future. Swift spent a great deal of time there

between December 2015 and October 2017, documenting how the island and its people were responding to climate change. This book is a chronicle of what he learned about the lives of Tangier Islanders and about their history, culture, and beliefs, set within the context of what is known about the likely fate of this vulnerable, isolated island.



**Chesapeake Requiem**  
Earl Swift  
Dey Street Books,  
2018. 448 pp.

Swift paints vivid portraits of both the natural environment and the individuals and institutions of this close-knit community, now fewer than 500 in number. Most residents depend on the fishery for their livelihood. Many share an ancestry on the island going back generations, and most share a deep Christian faith, a strict moral code, and a commitment to hard work in a challenging environ-

ment. As they observe rapid changes in the landscape around them, the islanders fear the loss of the land they call home, and they ask and pray for both government assistance and divine intervention to forestall what most scientists regard as virtually inevitable.

Swift's account of his time on the island is roughly chronological, interspersed with passages recounting the island's history and sections that cite the scientific literature about the ongoing

geologic evolution of the bay, predicted trends in sea-level rise, and the life cycle of the blue crab. He demonstrates a deep affection and admiration for the Tangier Islanders, who are often cantankerous but whom he also describes as "warm, loving, generous" and "remarkably resilient, hardworking, and courageous." As he illustrates in one of the more harrowing and moving episodes recounted in the book, they are even willing to risk their lives for one another.

One of the enduring mysteries documented in this book is how social and religious attitudes on Tangier Island and elsewhere have become so entwined with an overt refusal to believe the scientific evidence in support of the rapidly accelerating human impact on climate and sea-level rise. As Swift reveals, Tangier Islanders understand the effect of wind and waves on erosion but steadfastly maintain that human-induced climate change and sea-level rise are not real. Those he interviews believe that a government-funded seawall would save their island.

Swift clearly does believe the science. He points out that thousands of coastal communities in the United States (and many more around the world) either are or will soon be at risk and that it is both physically and economically impossible to save them all. He suggests that how we choose to respond to Tangier Island will say much about what we hold important.

Recent news reports suggest that Tangier Island may be in line for a government-funded coastal protection project (2). The previous administration awarded funding to relocate the "climate refugees" of Isle de Jean Charles in Louisiana (3). Looking to the future, these types of decisions will need to be based on an understanding of the rate and consequences of sea-level rise and of the cumulative costs and trade-offs that are involved when multiple communities are threatened (4).

This book offers a well-rounded portrait of a rural community both dependent on and threatened by its natural environment—and a description of a problem that unfortunately will become commonplace over the coming decades. ■

## REFERENCES

1. D. M. Schulte *et al.*, *Sci. Rep.* **5**, 17890 (2015).
2. T. Dietrich, *Daily Press*, 24 May 2018.
3. M. I. Stein, *Wired*, 25 January 2018.
4. J. Moore, L. Acker, "Recurrent flooding, sea level rise, and the relocation of at-risk communities: Case studies from the Commonwealth of Virginia" (Virginia Coastal Policy Center, 2018).

10.1126/science.aau3895

The reviewer is at the Department of Geography and Environmental Systems, University of Maryland, Baltimore County, Baltimore, MD 21250, USA. Email: miller@umbc.edu

## ENVIRONMENTAL TOXICOLOGY

# Our looming lead problem

Neglect, poor planning, and bad decisions led to Flint's water crisis. It could easily happen again.

By **Frederick Rowe Davis**

**O**n 25 April 2014, Flint Mayor Dayne Walling ceremoniously shut off a valve to the Detroit water supply and opened the flow of the Flint River into local homes and businesses. He marked the occasion by drinking from a glass filled with water from the new source. Eighteen months later, the water supply was switched back to Detroit. What occurred in between—the city's failure to control infrastructure corrosion, the deterioration of fresh water entering Flint residences, citizen complaints, government denials, elevated lead levels in children, and public outcry—would become the basis for a crisis that rose to national attention in 2015. *What the Eyes Don't See* by Mona Hanna-Attisha and *The Poisoned City* by Anna Clark are the first two books to document in detail this horrible disaster.

In *What the Eyes Don't See*, the Flint story unfolds from the strikingly personal perspective of Hanna-Attisha, the pediatrician and activist who documented elevated blood lead levels in the city's children. Born in England to Iraqi immigrants, Hanna-Attisha moved with her family to Flint when she was a child. Her brother, an attorney, has represented whistle-blowers in corporate fraud cases. A distant cousin was the first bacteriologist of Middle Eastern descent to work in the United States. A friend from high school became an environmental engineer at the U.S. Environmental Protection Agency. These relationships, we learn, bolster Hanna-Attisha's commitment to her community.

Yet, "Dr. Mona," as she is known, never strays too far from the stories of her young patients in Flint, whom she refers to as her "Flint kids." With names changed for privacy, the children's stories emphasize the importance of tap water to daily life and the long-term risks associated with lead exposure.

Hanna-Attisha's writing style is direct and passionate. Her descriptions of the events that unfolded and the people in-

involved reverberate with candor. She pulls no punches, directing some of her strongest criticism at Charles Kettering, whom she calls "Public Health Enemy #1."

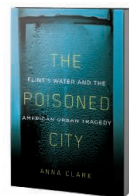
Head of research at General Motors (GM) from 1920 to 1947, Kettering was the driving force behind the world's embrace of leaded gasoline during the mid-20th century. Despite known alternatives, Kettering promoted tetraethyl lead, knowing GM could patent the additive. Ironically, and despite his role in the history of lead poisoning, there are tributes to Kettering all over Flint.

If Kettering is the villain, Alice Hamilton is Hanna-Attisha's heroine. A physician and research scientist during the early 20th century, Hamilton fought the use of lead in gasoline and advocated for safety

**What the Eyes Don't See**  
A Story of Crisis,  
Resistance, and Hope in  
an American City  
*Mona Hanna-Attisha*  
One World, 2018.  
378 pp.



**The Poisoned City**  
Flint's Water and the  
American Urban Tragedy  
*Anna Clark*  
Metropolitan Books,  
2018. 318 pp.



Lead is ubiquitous in American infrastructure. Minute amounts in water can accumulate to toxic levels, leading to cognitive deficits.

standards that would protect workers exposed to it from its effects. Hamilton's passionate fight clearly inspired Hanna-Attisha, who used ZIP codes to demonstrate that children living in areas that depended on water from the Flint river exhibited elevated blood lead levels.

*The Poisoned City* centers the story on Flint families and their advocacy on behalf of their children. Clark notes that Flint's water crisis did not result from corporate malfeasance nor natural disaster. Rather, she argues, it arose as a direct result of neglect (at times willful), poor planning, and bad decision-making (the latter including city officials' decision to violate the Safe Drinking Water Act, followed by months of delay and cover-up at all levels of government).

But the story of Flint is also one of resilience. Residents rose up and challenged the indifference and false palliatives offered by the authorities, who for many months proclaimed that the water supply in Flint was safe for drinking.

Clark provides context to the Flint disaster with informative sections on race in America, the rise and fall of cities, and environmental justice. She details events and developments as recent as the spring of 2018, including positive signs of reinvestment in Flint's social and economic infrastructure. (Hanna-Attisha provides similar details, often filtered through the experiences of immigrants to America.)

Both *The Poisoned City* and *What the Eyes Don't See* offer day-by-day (at times moment-by-moment) accounts of the inexorable development of the conditions that led to the Flint disaster. If Hollywood were to produce a film inspired by the crisis, one can imagine scenes being interspersed with graphics revealing specific dates and times. Both books would have benefited from a similar device, if only to underscore how quickly the crisis emerged.

The water crisis in Flint is profoundly worrisome: Numerous children suffered lead poisoning as a direct result of a bureaucratic focus on the fiscal rather than the social. With the huge amount of lead incorporated into the nation's infrastructure, many other communities are just a few poor decisions from a similar fate.

Double reviews prompt the question of which book to read. My recommendation is to read both. *What the Eyes Don't See* and *The Poisoned City* reverberate with crucial details, profound insights, and inspiring conclusions. This is a rare opportunity to read two superb—yet strikingly distinct—accounts of a national tragedy. ■

The reviewer is at the Department of History,  
Purdue University, West Lafayette, IN 47907, USA.  
Email: frdavis@purdue.edu

10.1126/science.aau3894





## LETTERS

UCI decided to remove Francisco Ayala's name from the science library after he resigned amid sexual harassment allegations.

Edited by **Jennifer Sills**

### **Harassment charges: Enough himpathy**

We are well into the #Metoo era, yet journalists and editors are still fixated on the harasser's fall from grace rather than the detrimental effect of sexual harassment on the victims and our society as a whole. The News story "Prominent geneticist out at UC Irvine after harassment finding" (M. Wadman, 29 June, <https://scim.ag/AyalaResignation>) reinforces a familiar toxic narrative: The accomplishments of the harasser hold more value to science than women's right to a safe workplace. This is now so commonplace that it has been dubbed "himpathy" (1).

In the News story, Wadman tells us all about the "eminent" professor, from his scientific accomplishments to his personal hobbies. He did "pioneering" and "ground-breaking" work, he donated money to the university, and he was president of AAAS (the publisher of *Science*). However, we do not hear about the pioneering work of the women he harassed at University of California, Irvine (UCI). From a graduate student to a tenured professor to an assistant dean, the News story reduced the women who demanded an end to his misconduct to complainers. We are told that Ayala was just being "European" and his actions were misunderstood; instead, the narrative should focus on the many women and careers that suffered from Ayala's actions. Wadman then chose to end the article by quoting an Ayala supporter who diminished the investigation.

The same himpathy sentiments return in the follow-up News In Depth story "Report details harassment by famed biologist" (M. Wadman, 27 July, p. 316). Words matter, and *Science* should wield its words and influence carefully. It is time to recognize that harassers have taken a substantial toll on the advancement of science. It is time to acknowledge that sexual harassment in all its nefarious forms puts an unquantifiable burden on the victims (many of whom are our colleagues). It's time to believe women.

**Jane Zelikova, Kelly Ramirez,\*  
Jewel Lipps, on behalf of 500 Women  
Scientists leadership**

500 Women Scientists, Boulder, CO 80303, USA.

\*Corresponding author.

Email: [k.ramirez@nioo.knaw.nl](mailto:k.ramirez@nioo.knaw.nl)

#### REFERENCE

1. K. Manne, *Down Girl* (Oxford Univ. Press, 2017).

10.1126/science.aau6858

### **Harassment charges: Journalists' role**

As a woman in science, I find the reporting of allegations against Francisco Ayala ("#Metoo complaints fell noted geneticist," M. Wadman, News In Brief, 6 July, p. 8) deeply troubling. The title implicates the "complaints" or "complainers" as responsible for the resignation, as opposed to the actions of the accused or the impartial sexual harassment investigation. The term "complaint" gives the impression that the alleged victims have minor grievances as opposed to serious reports of scientific misconduct. Not only is this title biased

against whistleblowers, it is also factually incorrect, as the first allegations of harassment were allegedly made 3 years before the resignation. I assert that the investigation, not the allegations, caused the accused to resign.

**Jessica Duffy**

Department of Biology, Midwestern State University,  
Wichita Falls, TX 76308, USA.

Email: [jlduffy1009@my.mwsu.edu](mailto:jlduffy1009@my.mwsu.edu)

10.1126/science.aau7641

### **Harassment charges: Injustice done?**

We are deeply concerned by the way in which our friend and colleague Professor Francisco Ayala has been forced to resign from the University of California, Irvine (UCI), after being accused of sexual harassment ("Prominent geneticist out at UC Irvine after harassment finding," M. Wadman, News, 29 June, <https://scim.ag/AyalaResignation>). The charges that have been raised against him have had appalling consequences.

Those of us who are well acquainted with Professor Ayala know that he is an honorable person, who throughout his career has treated his friends, co-workers, and students in a respectful, egalitarian way. His lifelong commitment to teaching, research, and outreach on biological evolution has won him worldwide recognition. He has been a generous benefactor to the University of California and throughout his fruitful career has opened new fields of biological research, promoted mutual respect and independence

between evolutionary studies and religious perspectives, played a key role in several major scientific organizations, and helped many Spanish-speaking female scholars and Hispanic students, in particular, both in the United States and throughout the world.

From the available information, it appears that the inquiry conducted by UCI lacked genuine due process, fairness, and full transparency. We urge UCI to acknowledge the possibility that its sanctions against Professor Ayala were enacted in haste and to reopen the case and investigate the matter more thoroughly. We understand the wish of both the institution and Professor Ayala not to unduly prolong this whole unhappy episode. It is equally important, however, that justice be done and be seen to be done. If carried out properly, UCI could help devise a more successful model for how institutions should deal with such situations in the future. Devising an improved procedure for these cases would earn everyone's gratitude.

**Andrés Moya and 62 additional authors\***

Institute of Integrative Systems Biology, University of València, 46010 València, Spain, and Foundation for the Promotion of Health and Biomedical Research of València Region (FISABIO), 46020 València, Spain. Email: andres.moya@uv.es

\*The full list of authors is available online.

**SUPPLEMENTARY MATERIALS**

Full list of authors

[www.sciencemag.org/content/361/6403/655.3/suppl/DC1](http://www.sciencemag.org/content/361/6403/655.3/suppl/DC1)

10.1126/science.aau7088

## *Harassment charges: Metoo but due process*

The resignation of eminent biologist Francisco J. Ayala amid charges of sexual harassment ("Prominent geneticist out at UC Irvine after harassment finding," M. Wadman, News, 29 June, <https://scim.ag/AyalaResignation>) has left Ayala's home campus deeply divided. The controversy highlights problems in the way universities currently address charges of sexual harassment. Because the University of California, Irvine (UCI), failed to post easily accessible guidelines on how the Ayala case was handled, especially how and by whom final decisions were made, many UCI faculty are concerned that the university overreached, imposing a punishment not commensurate with the specific charges of wrongdoing. A zero-tolerance policy for sexual harassment should include clearly stated procedures to protect due process and ensure proportional responses to wrongdoing.

I will not engage in the ugly nastiness of "he said, she said" as we debate the

veracity of accusations described in salacious detail ("Report details harassment by famed biologist," M. Wadman, News In Depth, 27 July, p. 316). These are human tragedies evolving in front of us. Victims strive to regain lost self-esteem, along with justice. The accused wrestle with shock, forced to confront their own cluelessness at shifting mores, and everyone realizes how vulnerable we all are, including administrators struggling to figure out how best to fairly confront sexual harassment and gender equality. We are all at risk when goodwill and communications break down.

To fairly and equitably strike a balance between complacency and overreaction, I recommend three procedures to help achieve the notoriously difficult goal of changing attitudes toward gender. First, we must do more to foster a university climate that takes seriously and protects potential targets, be they male or female. Part of that is recognizing that even subtle forms of verbal behavior—i.e., jokes or comments about appearance—make some women and minorities uncomfortable. Yet resulting deferential treatment can contribute to a climate in which women or minorities are treated differently. Eventually, this subtle, differential treatment can foster continuing inequality. Second, we must insist that legally correct policies and procedures are followed when charges of sexual harassment are made. These procedures need to be transparent and the same for all faculty. Policies should be written in prose that even nonlawyers can understand and posted in obvious places, with regularly scheduled discussion forums designed to help educate all members of the university community. Third, the punishment must fit the crime. If tasteless, off-color jokes and the kind of ambiguous "unwanted touching" of which Ayala was accused warrant his public humiliation, what do we do with more serious charges of sexual harassment? And why is Hana Ayala punished for her husband's acts by having her name removed from gifts to UCI?

The #Metoo movement has done a great service in opening up an area too long taboo. But the failure to follow clearly established, fairly administered, and transparent procedures can too easily produce witch hunts that cast doubts on legitimate charges of sexual harassment. This will set back the move toward gender equality, in the academy and in society at large.

**Kristen Renwick Monroe**

Department of Political Science, University of California, Irvine, Irvine, CA 92697, USA.  
Email: [krmonroe@uci.edu](mailto:krmonroe@uci.edu)

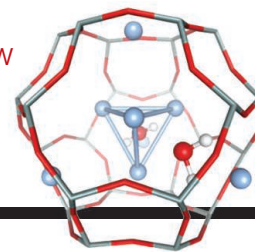
10.1126/science.aau7986



# RESEARCH

Silver's bright glow  
in captivity

Grandjean et al., p. 686



IN SCIENCE JOURNALS

Edited by **Caroline Ash**



## ACTIVE MATTER

### When fewer workers are more efficient

**A** narrow passageway can easily become clogged or jammed if too much traffic tries to enter at once or there is competition between the flow of traffic in each direction. Aguilar *et al.* studied the collective excavation observed when ants build their nests. Because of the unequal workload distribution, the optimal excavation rate is achieved when a part of the ant collective is inactive. Numerical simulations and the behavior of robotic ants mimic the behavior of the colony. —MSL

*Science*, this issue p. 672

**Ants optimize their excavation rate in narrow tunnels when only some are active.**

## PLANT SCIENCE

### Pest subverts host plant's foraging

Plants need iron as a micronutrient, and they extract it from the rhizosphere by secreting chelating agents. Insect pests, such as the western corn rootworm, which annually cause millions of dollars' worth of lost yield, need iron, too. Hu *et al.* show that the rootworm exploits the plant's own iron-foraging system to detect its host and to seize iron for itself (see the Perspective by Kliebenstein). Plants produce benzoxazinoid compounds not only as a defense against many insects but also as iron chelators. Rootworm larvae are not harmed by benzoxazinoids; instead, they take advantage of their presence as a signal that food is near and

of their properties as an iron chelator. —PJH

*Science*, this issue p. 668;  
see also p. 642

## CLIMATE

### A shrinking marine refrigerator

Low subtropical marine clouds scatter solar radiation back to space and thereby cool the climate system. Most work on understanding changes in the coverage of these types of clouds has focused on the effects of sea surface temperatures or on aerosols. Yuter *et al.* show that dynamic effects due to atmospheric gravity waves are responsible for the rapid clearing of large areas of these clouds. This phenomenon also has

implications for marine ecology and biogeochemistry. —HJS

*Science*, this issue p. 697

## SOLID-STATE PHYSICS

### Controlling two- dimensional twist

In heterostructures assembled from two-dimensional materials such as graphene, electron tunneling between layers varies strongly with the rotation angle between the crystal lattices. Usually, the twist angle between layers is fixed after assembly. Ribeiro-Palau *et al.* encapsulated graphene with boron nitride, but the top boron nitride flake was shaped so that an atomic force microscope tip could push on it to vary the twist angle by as little as 0.2°.

They observed variations with twist angle in properties such as the charge neutrality point, which would be difficult to observe in static rotated structures. —PDS

*Science*, this issue p. 690

## IMMUNOLOGY

### Liquid droplets step on the cGAS

Spontaneous partitioning of a homogeneous solution of molecules, or liquid-phase separation, underlies the formation of cellular bodies from P granules to nucleoli. Essentially, dense-phase liquid droplets act like cellular compartments. Du *et al.* show that DNA binding to its cytoplasmic sensor, cyclic GMP–AMP synthase

CREDITS: (PHOTO) BERT PLUS/INIS/MINDEN PICTURES; (GRAPHIC) D. GRANDJEAN ET AL.

(cGAS), results in liquid droplets containing activated cGAS (see the Perspective by Ablasser). This phenomenon occurs through multivalent interactions, augmented by zinc, between DNA binding domains on cGAS and DNA in a length-dependent manner. Binding triggers a switchlike reaction that concentrates the enzyme and reactants to enhance STING-dependent interferon responses. —STS

*Science*, this issue p. 704;  
see also p. 646

## RNA METABOLISM

### A tale of not-so-pure tails

The poly(A) tail of mRNA has been thought to be a pure stretch of adenosine nucleotides with little informational content except for length. Lim *et al.* identified enzymes that can decorate poly(A) tails with non-A nucleotides. The noncanonical poly(A) polymerases, TENT4A and TENT4B, incorporate intermittent non-A residues (G, U, or C) with a preference for guanosine, which results in a heterogeneous poly(A) tail. Deadenylation trim poly(A) tails to initiate mRNA degradation but stall at the non-A residues. In effect, the not-so-pure tail stabilizes mRNAs by slowing down deadenylation. —BAP

*Science*, this issue p. 701

## SOCIAL ROBOTICS

### Can robots peer-pressure children?

Most of us know the classic peer-pressure question: "If everyone jumped off a bridge, would you?" Vollmer *et al.* investigated whether robots can have the same effect on humans as our conspecifics do. Child or adult participants were asked to determine which of three lines on a screen was the same as a reference line, but there was a secret twist: Three confederates seated with the participant all gave the same wrong answer. Adults were more likely to go along with the group's wrong answer when the confederates were all humans but resisted

the social pressure from a group of robots. Children, however, were susceptible to robot peer pressure, too, which raises the question of what safeguards are needed when young or vulnerable populations interact with automata. —RLK

*Sci. Robot.* **3**, eaat7111 (2018).

## ENVIRONMENTAL STUDIES

### Forests improve dietary diversity

Micronutrient deficiency affects nearly two billion people worldwide and particularly affects children in developing countries. Analyzing data on children's diets in more than 43,000 households across 27 developing countries, Rasolofson *et al.* show that access to local forests, and the foods they provide, directly improves dietary diversity and mitigates nutrient deficiencies comparably to nutrition-sensitive agricultural programs. The magnitude of this connection between nutrition and forests is greatest in communities with lower levels of development and implies that nutrition deficiencies in developing communities may be driven by the loss or poor management of local forests. —AC

*Sci. Adv.* 10.1126/sciadv.aat2853 (2018).

## HIGH-PRESSURE PHYSICS

### Laser-shocking deuterium into metal

The conditions in which hydrogen disassociates and becomes an atomic metal occur in high-energy-density environments, such as the interiors of giant planets and nuclear explosions. Celliers *et al.* trained 168 lasers on deuterium samples at the National Ignition Facility to measure the pressure and temperature conditions of the hydrogen transition. Careful optical measurements led to the addition of four new points on the phase diagram, consistent with static estimates and theoretical calculations. —BG

*Science*, this issue p. 677

## IN OTHER JOURNALS

Edited by **Caroline Ash**  
and **Jesse Smith**



## HIV

### Viral genomes behind HIV remission

HIV-positive patients are often dependent on continuous antiretroviral therapy (ART) to prevent AIDS progression (such patients are called noncontrollers). A small proportion of individuals, called HIV post-treatment controllers, sustain HIV remission after receiving short-term ART. To understand the differences between these types of patients, Sharaf *et al.* sequenced HIV genomes from plasma samples. Before the interruption of ART, posttreatment controllers had a lower HIV genome reservoir size than noncontrollers by a factor of 7. Posttreatment controllers also had fewer defective HIV genome copy numbers and showed heightened HIV-specific immune responses and slower viral rebound after interruption of ART. Analysis of HIV genome characteristics could provide important information for designing an individual's

optimal treatment plan. —MY  
*J. Clin. Invest.* 10.1172/JCI120549 (2018).

## BIOASSAYS

### Detection by protection

DNA aptamers are oligonucleotides that bind with high specificity to targets ranging from metal ions to whole proteins. Binding of the desired target can introduce structural changes in the DNA and thus signal the concentration of the target. Canoura *et al.* designed a DNA aptamer-based bioassay that can determine the concentration of multiple small molecules at once. If no target is present, enzymes degrade the DNA aptamer, which is unprotected on its own. Binding of the aptamer target locks the DNA in a stable conformation, protecting it from degradation and enabling readout through a second, labeled DNA. The assay was used to detect ATP and cocaine in a complex sample simultaneously. —MAF

*J. Am. Chem. Soc.* **140**, 9961 (2018).





## NEUROSCIENCE

### From the concrete to the abstract

**E**arly processing in the brain tends to simply represent external information from the sensory organs. However, to behave appropriately, organisms need to structure this input into meaningful categories and concepts. Brincat *et al.* recorded from multiple regions along the processing hierarchy in monkey brains while the animals performed sophisticated categorization tasks. Lower-level areas conveyed strong information about sensory stimuli within their preferred domains but indicated hardly any abstraction beyond the raw sensory inputs. In contrast, the prefrontal cortex, despite containing relatively weak information overall, showed strongly abstracted, task-relevant coding. Intermediate areas exhibited mixed representations, with partial categorical coding occurring in some domains but not others. Representational transformation thus happens gradually, across multiple cortical processing steps, rather than in a discrete, all-or-nothing fashion. —PRS

*Proc. Natl. Acad. Sci. U.S.A.* 10.1073/pnas.1717075115 (2018).

**Monkeys use sophisticated categorization strategies—for example, while selecting what to eat.**

## PALEOCLIMATOLOGY

### Time is the key

How does one construct an accurate time scale for ocean sediments? One way is to measure their carbon-14 content and use that radioactive clock to date them. This approach has pitfalls, though, because the carbon fixed by organisms at the sea surface that is later locked up in sediments may itself have been old (i.e., the sediment clock may not have started at time zero). Zhao and Keigwin constructed a time scale for sediments from the eastern equatorial Pacific by dating twigs found within them, thereby eliminating the problem of correcting for surface reservoir effects. Ventilation ages of the thermocline and intermediate waters there during the Last Glacial Maximum and deglaciation were similar to those of today, contradicting results from a number of previous studies. —HJS

*Nat. Commun.* **9**, 3077 (2018).

## ORGANISMAL BIOLOGY

### Up through the atmosphere

Surprisingly, a major component of the diet of swifts is spiders. Swifts only stop flying when they are breeding, so how do the spiders on which they feed become airborne? Morley and Robert discovered that spiders are sensitive to electric fields, so that they respond to the atmospheric potential gradient. The electric field mechanically activates hairs (trichobothria) on the spiders' legs. Formation of a vertical electric field stimulates a

spider standing on top of a plant to release fine threads of silk and let go. The electric field strength of pointed leaves can achieve tens of kilowatts per meter in mildly unsettled weather, and the atmospheric potential gradient provides enough coulomb force to lift a spider into the air without any help from passing breezes. The next thing to work out is what the advantages are to spiders to ascending sometimes thousands of meters into the atmosphere—it isn't just to offer themselves as tidbits to swifts. —CA

*Curr. Biol.* **28**, 2324 (2018).



**Spiders stand on prominences, emit strands of silk, and take to the air.**

## METABOLIC DISEASE

### Overfed mice got rhythm

Metabolic health is intertwined with the body's circadian clock, but the mechanisms underlying this connection are not fully understood. Studying mice, Guan *et al.* examined the effect of diet-induced obesity on circadian expression of genes in the liver. They found that obesity intensifies and synchronizes the liver circadian rhythms of enhancers that control the expression of two transcription factors with opposing effects on lipid metabolism. The result is enhanced synthesis and oxidation of fatty acids. Drugs activating one of these transcription factors, PPAR $\alpha$ , are already used clinically to reduce lipid levels. Interestingly, in obese mice, these drugs were most effective when delivered at peak PPAR $\alpha$  expression, indicating that their efficacy in humans might be optimized by administration at certain times of day. —PAK

*Cell* **174**, 831 (2018).

## ORGANIC CHEMISTRY

### All together now

Displacement reactions at saturated carbons often proceed by a concerted mechanism: A new bond forms at the same time as the old bond breaks. In aromatic rings, Meisenheimer complexes have been characterized with the incoming group bound before the leaving group departs, so a two-step mechanism is the default assumption. Kwan *et al.* now report that this assumption usually does not hold. They measured carbon kinetic isotope effects of several nucleophilic aromatic substitutions by using a sensitive  $^{19}\text{F}$  nuclear magnetic resonance technique; they then benchmarked theoretical methods for a much broader virtual study. The net result was that 83% of 120 simulated reactions were concerted. —JSY

*Nat. Chem.* 10.1038/s41557-018-0079-7 (2018).

## ALSO IN SCIENCE JOURNALS

Edited by **Caroline Ash****OPTICAL IMAGING**  
**More to imaging than meets the eye**

Traditional imaging techniques involve peering down a lens and collecting as much light from the target scene as possible. That requirement can set limits on what can be seen. Altmann *et al.* review some of the most recent developments in the field of computational imaging, including full three-dimensional imaging of scenes that are hidden from direct view (e.g., around a corner or behind an obstacle). High-resolution imaging can be achieved with a single-pixel detector at wavelengths for which no cameras currently exist. Such advances will lead to the development of cameras that can see through fog or inside the human body. —ISO

*Science*, this issue p. 660**WHEAT GENOME**  
**Insights from the annotated wheat genome**

Wheat is one of the major sources of food for much of the world. However, because bread wheat's genome is a large hybrid mix of three separate subgenomes, it has been difficult to produce a high-quality reference sequence. Using recent advances in sequencing, the International Wheat Genome Sequencing Consortium presents an annotated reference genome with a detailed analysis of gene content among subgenomes and the structural organization for all the chromosomes. Examples of quantitative trait mapping and CRISPR-based genome modification show the potential for using this genome in agricultural research and breeding. Ramírez-González *et al.* exploited the fruits of this endeavor to identify tissue-specific biased gene expression and coexpression networks during development and exposure to stress. These

resources will accelerate our understanding of the genetic basis of bread wheat. —LMZ

*Science*, this issue p. 662**DEVELOPMENTAL SIGNALING**  
**How Wnt ligands achieve specificity**

Wnt signaling is essential for development, tissue homeostasis, and disease. The 19 members of the Wnt family interact promiscuously with the 10 Frizzled receptors, raising the question of how ligand-specific discrimination is achieved in a biological context. Eubelen *et al.* used experiments in zebrafish to show that cells are equipped with decoding modules that bind Wnt with high specificity and trigger signal amplification via their recruitment into higher-order Frizzled signalosomes (see the Perspective by Kim and Goentoro). Thus, distinct Wnt ligand-receptor pairs can be targeted specifically for therapeutic purposes. —BAP

*Science*, this issue p. 663;  
see also p. 643**MEMBRANES**  
**Spraying makes it smoother**

Commercial reverse osmosis processes for water desalination use membranes made by the polymerization of polyamide at the oil/water interface. Chowdhury *et al.* show that thinner, smoother membranes can be made with an electrospray technique. Using high voltage, the two precursors are finely sprayed onto a substrate and polymerize on contact. The composition of the resulting membrane can be tuned on the basis of the proportion of the two components. At optimum conditions, the membranes appear to be better at desalination than current commercial reverse osmosis membranes. —MSL

*Science*, this issue p. 682**PHOTOLUMINESCENCE**  
**Unmasking the glow of silver clusters**

Small silver clusters stabilized by organic materials or inorganic surfaces can exhibit bright photoluminescence, but the origin of this effect has been difficult to establish, in part because the materials are heterogeneous and contain many larger but inactive clusters. Grandjean *et al.* studied silver clusters in zeolites, using x-ray excited optical luminescence to monitor only the emissive structures (see the Perspective by Quintanilla and Liz-Marzán). Aided by theoretical calculations, they identified the electronic states of four-atom silver clusters bound with water molecules that produce bright green emission—thus identifying candidate materials for application in lighting, imaging, and therapeutics. —PDS

*Science*, this issue p. 686;  
see also p. 645**ORGANIC CHEMISTRY**  
**Radically transforming light hydrocarbons**

The methane, ethane, and propane in natural gas are mostly inert under ambient conditions. Mainly they are burned to produce heat. Hu *et al.* show that a simple cerium salt paired with an alcohol can catalytically transform these and other simple hydrocarbons into reactive radicals at room temperature (see the Perspective by Kanai). The reactions rely on light to photolytically cleave cerium alkoxide bonds, producing alkoxy radicals that strip H atoms from the hydrocarbons and regenerate the alcohol. The resultant alkyl radicals readily add to azo compounds, olefins, and aromatics. —JSY

*Science*, this issue p. 668;  
see also p. 647**ORGANIC CHEMISTRY**  
**Four varieties of carbonyl sandwich**

Compounds with adjacent carbons sandwiched between two carbonyl (C=O) centers turn up frequently in organic chemistry. When these central carbons each have a substituent, there are four possible mutual geometries, all with potentially distinct biochemical properties. Kaldre *et al.* present a single method to access each stereoisomer individually. The outcome depends on the straightforwardly tunable configuration of a sulfoxide group in a precursor, which guides a rearrangement. The versatility of the method should facilitate selective access to 1,4-dicarbonyl motifs in pharmaceutical research. —JSY

*Science*, this issue p. 664**TRANSLATIONAL CONTROL****Fragile X and fragile translation in flies**

Mutations in the fragile X mental retardation 1 (*FMR1*) gene underlie fragile X syndrome and fragile X-associated primary ovarian insufficiency, which are prominent intellectual disability and reproductive disorders, respectively. *FMR1* is thought to reduce protein synthesis (translation) at synapses. In *Drosophila* oocytes, Greenblatt and Spradling found that *Fmr1* loss leads to oocytes that generate embryos exhibiting neural defects (see the Perspective by Aryal and Klann). Ribosome profiling of oocytes identified a specific role for *FMR1* in enhancing the translation of large proteins, including many associated with autism. *FMR1* seems to help maintain translation of large mRNAs that otherwise condense into inactive ribonucleoprotein particles. This mechanism may underlie other causes of autism and mental dysfunction. —BAP

*Science*, this issue p. 709;  
see also p. 648



## LIVER DISEASE

### Setting liver regeneration free

The mammalian liver has considerable regenerative capacity. After acute severe injury, as occurs in acetaminophen (paracetamol or Tylenol) poisoning, liver regeneration and recovery may fail. Bird *et al.* show that senescence, which is classically associated with aging and carcinogenesis, prevents liver cell proliferation and reduces the liver's regenerative capacity after acute injury. Senescence spreads from cell to cell through signaling by transforming growth factor- $\beta$  (TGF $\beta$ ). When TGF $\beta$  signaling was blocked during acetaminophen poisoning in mice, senescence was impeded, regeneration accelerated, and mouse survival increased—thus presenting an attractive therapeutic approach for liver regeneration. —OMS

*Sci. Transl. Med.* **10**, eaan1230 (2018).

## LYMPHATICS

### Setting the stage for attack

There is growing recognition of the importance of stromal cells in shaping immune organs and immune responses. In classical secondary lymphoid organs (SLOs) such as lymph nodes, tonsils, and Peyer's patches, fibroblastic reticular cells (FRCs) play an integral part in immune responses. Nonclassical SLOs, including fat-associated lymphoid clusters (FALCs), also have important roles in systemic immunity. Perez-Shibayama *et al.* report that FRCs in FALCs are both organizational and immunomodulatory. —AB

*Sci. Immunol.* **3**, eaar4539 (2018).

## HORMONE SIGNALING

### How progesterone stimulates cilia

The transport of eggs depends on the beating of the cilia of the cells lining the oviduct. Cilia beating is a Ca<sup>2+</sup>

signaling-dependent process. Jung *et al.* investigated how the hormone progesterone affects Ca<sup>2+</sup> signaling in mouse ciliated oviduct cells. The hormone increased intracellular Ca<sup>2+</sup> concentrations in a process requiring the cation channel TRPV4 and the stepwise activation of GABA<sub>A</sub> and GABA<sub>B</sub> receptors. These receptors associated with each other in response to progesterone and GABAergic agonists. —JFF

*Sci. Signal.* **11**, eaam6558 (2018).

## REVIEW SUMMARY

## OPTICAL IMAGING

# Quantum-inspired computational imaging

Yoann Altmann, Stephen McLaughlin, Miles J. Padgett, Vivek K Goyal, Alfred O. Hero, Daniele Faccio\*

**BACKGROUND:** Imaging technologies, which extend human vision capabilities, are such a natural part of our current everyday experience that we often take them for granted. However, the ability to capture images with new kinds of sensing devices that allow us to see more than what can be seen by the unaided eye has a relatively recent history.

In the early 1800s, the first ever photograph was taken: an unassuming picture that required days of exposure to obtain a very grainy image. In the late 1800s, a photograph was used for the first time to see the movement of a running horse that the human eye alone could not see. In the following years, photography played a pivotal role in recording human history, ranging from influencing the creation of the first national parks in the United States all the way to documenting NASA's Apollo 11 mission to put a man on the Moon. In the 1990s, roughly

10 billion photographs were taken per year. Facilitated by the explosion in internet usage since the 2000s, this year we will approach 2 trillion images per year—nearly 1000 images for every person on the planet. This upsurge is enabled by considerable advances in sensing and data storage and communication. At the same time, it is driving the desire for imaging technology that can further exceed the capabilities of human vision and incorporate higher-level aspects of visual processing.

**ADVANCES:** Beyond consumer products, research labs are producing new forms of imaging that look quite different from anything we were used to and, in some cases, do not resemble cameras at all.

Light is typically detected at relatively high intensities, in the spectral range and with frame rates comfortable to the human eye. However,

emerging technologies are now relying on sensors that can detect just one single photon, the smallest quantum out of which light is made. These detectors provide a “click,” just like a Geiger detector that clicks in the presence of radioactivity. We have now learned to use these “click” detectors to make cameras that have enhanced properties and applications. For example, videos can be created at a trillion frames per second, making a billion-fold jump in speed with respect to standard high-speed cameras. These frame rates

## ON OUR WEBSITE

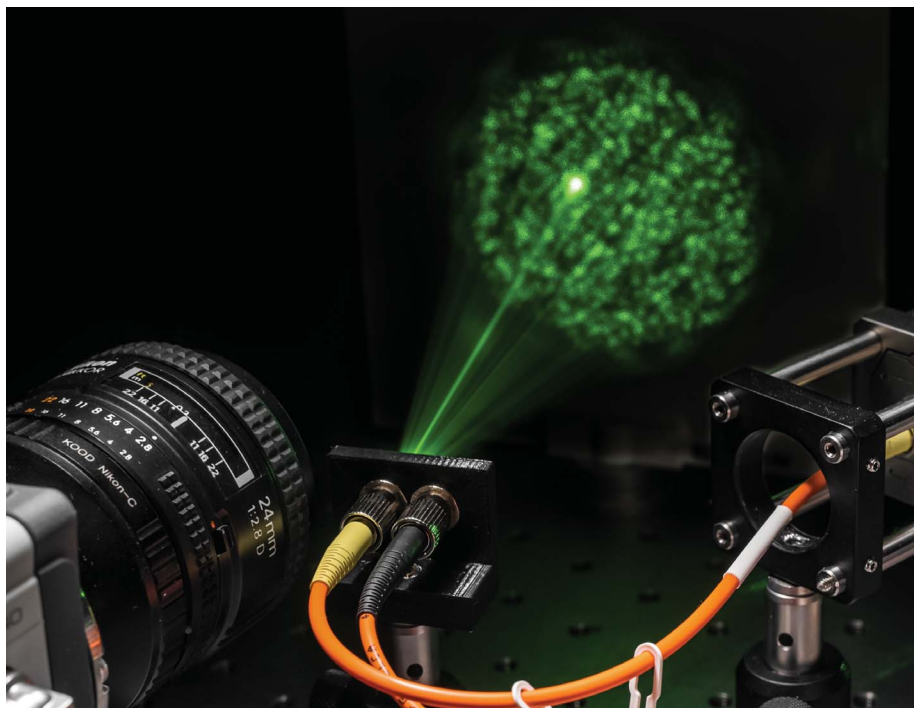
Read the full article at <http://dx.doi.org/10.1126/science.aat2298>

are sufficient, for example, to freeze light in motion in the same way that previous photography techniques were able to freeze the motion of a bullet—although light travels a billion times faster than a supersonic bullet. By fusing this high temporal resolution together with single-photon sensitivity and advanced computational analysis techniques, a new generation of imaging devices is emerging, together with an unprecedented technological leap forward and new imaging applications that were previously difficult to imagine.

For example, full three-dimensional (3D) images can be taken of a scene that is hidden behind a wall, the location of a person or car can be precisely tracked from behind a corner, or images can be obtained from a few photons transmitted directly through an opaque material. Inspired by quantum techniques, it is also possible to create cameras that have just one pixel or that combine information from multiple sensors, providing images with 3D and spectral information that was not otherwise possible to obtain.

**OUTLOOK:** Quantum-inspired imaging techniques combined with computational approaches and artificial intelligence are changing our perspective of what constitutes an image and what can or cannot be imaged. Steady progress is being made in the direction of building cameras that can see through fog or directly inside the human body with groundbreaking potential for self-driving cars and medical diagnosis. Other cameras are being developed that can form 3D images from information with less than one photon per pixel. Single-photon cameras have already made their way into widely sold smartphones where they are currently used for more mundane purposes such as focusing the camera lens or detecting whether the phone is being held close to one's ear. This technology is already out of the research laboratories and is on the way to delivering fascinating imaging systems. ■

The list of author affiliations is available in the full article online.  
\*Corresponding author. Email: [daniele.faccio@glasgow.ac.uk](mailto:daniele.faccio@glasgow.ac.uk)  
Cite this article as Y. Altmann et al., *Science* 361, eaat2298 (2018). DOI: 10.1126/science.aat2298



Quantum-based imaging systems are being developed to image through opaque media (e.g., fog or human tissue) that scatter light in all directions.

PHOTO: KEVIN J. MITCHELL



## REVIEW

## OPTICAL IMAGING

# Quantum-inspired computational imaging

Yoann Altmann<sup>1</sup>, Stephen McLaughlin<sup>1</sup>, Miles J. Padgett<sup>2</sup>, Vivek K Goyal<sup>3</sup>, Alfred O. Hero<sup>4</sup>, Daniele Faccio<sup>2\*</sup>

Computational imaging combines measurement and computational methods with the aim of forming images even when the measurement conditions are weak, few in number, or highly indirect. The recent surge in quantum-inspired imaging sensors, together with a new wave of algorithms allowing on-chip, scalable and robust data processing, has induced an increase of activity with notable results in the domain of low-light flux imaging and sensing. We provide an overview of the major challenges encountered in low-illumination (e.g., ultrafast) imaging and how these problems have recently been addressed for imaging applications in extreme conditions. These methods provide examples of the future imaging solutions to be developed, for which the best results are expected to arise from an efficient codesign of the sensors and data analysis tools.

Computational imaging is the fusion of computational methods and imaging techniques with the aim of producing better images, where “better” has a multiplicity of meanings. The development of new imaging sensors and, in particular, instruments with single-photon sensitivity, combined with a new wave of computational algorithms, data handling capability, and deep learning, has resulted in a surge of activity in this field.

One clear trend is a shift away from increasing the number of megapixels and toward fusing camera data with computational processing and, if anything, decreasing the number of pixels, potentially to a single pixel. The incoming data may therefore not actually look like an image in the conventional sense but are transformed into one after a series of computational steps and/or modeling of how the light travels through the scene or the camera. This additional layer of computational processing frees us from the chains of conventional imaging techniques and removes many limitations in our imaging capability.

We briefly describe some of the most recent developments in the field, including full three-dimensional (3D) imaging of scenes that are hidden (e.g., around a corner or behind an obstacle), high-resolution imaging with a single-pixel detector at wavelengths for which no cameras exist, cameras that can see through fog or inside the human body, and cameras that mimic the human eye by creating detail only in areas of interest. We will also discuss how multispectral

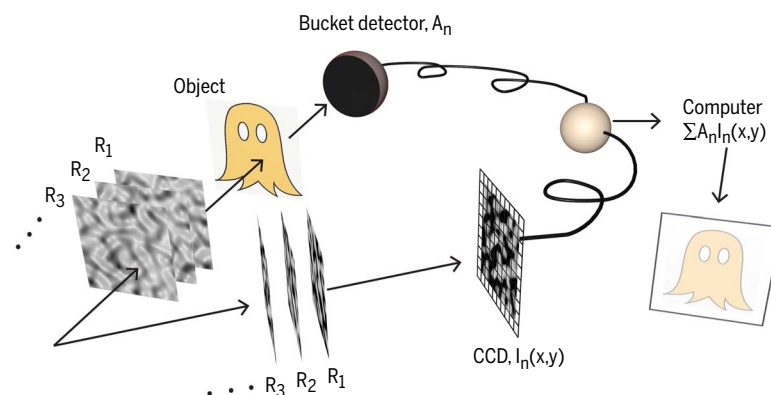
imaging with single-photon detectors can improve 3D reconstruction and provide richer information about a scene.

We discuss how single-photon detection technologies are transforming imaging capabilities with single-photon-sensitive cameras that can take pictures at the lowest light levels and with the ability to create videos reaching a trillion frames per second. This improvement has enabled the capture of images of light beams traveling across a scene and provided opportunities to observe image distortions and peculiar relativistic temporal inversion effects that are due to the finite speed of light. The ability to capture light in flight also underpins some of the applications mentioned above—for example, the ability to view a 3D scene from around a corner. Probabilistic modeling of the particlelike nature

of light when using single-photon detectors has stimulated the birth of new computational techniques such as “first-photon imaging,” which hints at the ultimate limits of information to be gained from detecting just one photon.

## Single-pixel and ghost imaging

Although most imaging techniques that have emerged recently are based on classical detectors and cameras, some of these approaches have been inspired by or have a tight connection with similar ideas in quantum imaging. A prime example is ghost imaging (Fig. 1) (1), originally thought to be based purely on quantum principles but now recognized as being dependent on spatial correlations that can arise from both quantum and classical light (2). The realization that this technique does not require quantum light led to a merging of the fields of computational ghost imaging (3) and work on single-pixel cameras (4), as well as to an overall increase of activity in the field. In its quantum version, ghost imaging refers to the use of parametric down-conversion to create pairs of photons with correlated positions. If we detect the position of one photon with a standard camera and illuminate an object with the other position-correlated photon, it is sufficient to detect only the reflectance or transmittance of the object with a single-pixel detector—i.e., to measure the correlation count between the beams to then reconstruct a full image by repeating the measurement with many different photon pairs (each of which will be randomly distributed because of the random nature of the correlated photon pair generation process) (5, 6). It is now acknowledged that the quantum properties of the correlated photons play no role in the image reconstruction process: Thermal light split into two beams using a beam splitter can be used equally effectively, albeit at a higher photon flux (7). Rather than using a beam splitter, it is possible to use a spatial light modulator to create a pattern where the copy is simply the computer memory. This approach



**Fig. 1. Ghost imaging.** Random spatial patterns,  $R_n$ , illuminate an object and only the total transmitted (or reflected) light,  $A_n$ , is measured. This intensity reading is then computationally combined with the information of the pattern,  $I_n(x, y)$  (either measured separately or known if generated by a computer), to form an image of the object. CCD, charge-coupled device.

<sup>1</sup>School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, UK. <sup>2</sup>School of Physics and Astronomy, University of Glasgow, Glasgow, UK. <sup>3</sup>Department of Electrical and Computer Engineering, Boston University, Boston, MA, USA. <sup>4</sup>Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA.

\*Corresponding author. Email: [daniele.faccio@glasgow.ac.uk](mailto:daniele.faccio@glasgow.ac.uk)

therefore no longer requires a camera of any kind in the setup: The computer-generated pattern is already known and the image,  $I$ , can be retrieved by multiplying the single-pixel readout,  $a_i$ , with the corresponding pattern,  $H_i$ , and then summing over all patterns, i.e.,  $I = \sum_i a_i H_i$ . This opens the route to so-called compressed single-pixel imaging, in which assumptions about the spatial correlations of the image enable patterns to be judiciously chosen to require far fewer patterns than the final number of image pixels, with compression factors as high as 80 to 90%. This concept is not dissimilar from standard JPEG compression, which assumes that typical images are concentrated in their spatial frequencies, with the difference that now the compression is applied at the image acquisition stage. By this compression, single-pixel imaging is therefore transformed from a slow, relatively inefficient process into a highly efficient imaging technique that can operate at video frame rates in full color (8). More recent developments include extensions to lensless imaging (9) and to full 3D images for which depth information is obtained by also using time-of-flight information (10, 11)—i.e., in addition to object reflectivity, the imaging system also estimates the light travel distance,  $d$ , from the temporal shift,  $\tau$ , of the detected signal, as the two are related by the speed of light,  $d = c\tau$ , in free space, where  $c$  is the speed of light. In general, this single-pixel technique suffers from having low resolution and providing poor-quality images even when compared with a cell-phone camera. This limitation may be partly overcome by taking inspiration from nature and implementing computational algorithms so that the system increases the density of the projected spatial patterns only in areas of interest, therefore increasing the spatial resolution in regions where it is needed and leaving the surrounding areas relatively less defined (12). This is just one example of computational techniques being combined with detection technology to provide more efficient sensing solutions. Another example is the first-photon imaging approach that emerged from a codesign of hardware and computational algorithms, built around the concept of single-photon detection.

### First-photon imaging

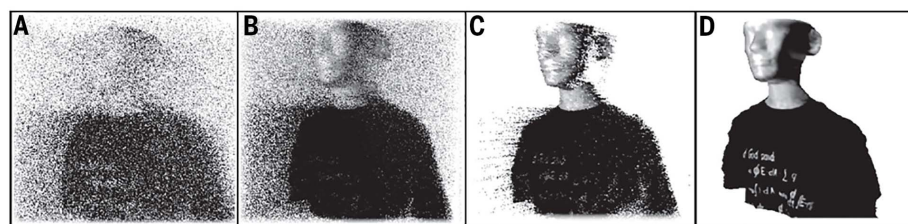
An important legacy of recent interest in the field of quantum information science is the development of a series of detector technologies for single photons. The workhorse for most laboratories is the single-photon avalanche diode (SPAD). SPADs are, in essence, semiconductor diodes that are reverse-biased beyond their breakdown threshold: A single photon (or even a thermally generated charge in the diode) is sufficient to lead to the rapid charge multiplication process (or avalanche) that creates a spike in the output current. A quenching mechanism stops the avalanche process before the diode is irreversibly damaged, leading also to a dead time during which the diode is insensitive to incident photons before being reactivated. The particle-like nature of a photon is revealed through the

very short burst in the SPAD output current that can then be very precisely timed when a reference signal is also provided. The ability to precisely detect the photon arrival time can be used for long-distance, high-precision light detection and ranging (LIDAR). A distant object is illuminated with a pointlike pulsed laser beam. Each outgoing pulse starts a counter, which is then stopped at time  $\tau$  when a return photon is detected; accounting for the two directions of the light travel, the distance of the illuminated object is simply  $c\tau/2$ . Scanning the scene using this time-correlated single-photon counting (TCSPC) technique can therefore provide a full 3D image (or depth image) of the scene (13–15). However, TCSPC-based imaging can require very long acquisition times, in particular when photons return to the detector at a low rate. Conventional processing techniques require (i) operation in the photon-starved regime (i.e., 10% or less of the outgoing laser pulses should give rise to a detected return photon so that bias from detector dead times is negligible) and (ii) measurement over many illumination repetition

background and the particlelike nature of the photons and their detection) built into the information used to retrieve high-quality 3D images. Similar extreme photon efficiency can be achieved with a fixed dwell time at each scene position (18), and principled statistical techniques for adapting the local spatial resolution to characteristics of the data enable background noise 25 times as strong as the back-reflected signal to be mitigated (19). Additional performance improvements have been obtained with deep learning (20). Using an array of SPADs parallelizes the data acquisition and thus can increase imaging speed, though an array has coarser time resolution, translating to coarsened longitudinal distance measurements (21). Methods for arrays can also be highly photon efficient (22).

### Non-line-of-sight imaging

Photon counting has strongly affected the field of non-line-of-sight (NLOS) imaging—i.e., the imaging of objects that are, for example, hidden behind a wall, corner, or obstacle (23–32). Access



**Fig. 2. First-photon imaging.** (A) Each photon detection can be mapped to a 3D position, which is often far from correct because half of the detections are due to background light. The number of illumination pulses until a photon is detected is inversely proportional to an initial estimate of reflectivity. (B) Exploiting piecewise smoothness yields improved reflectivity estimates. (C) Approximate noise censoring removes most detections due to background light. (D) The final estimate exploits piecewise smoothness of depth. [Adapted from figure 2 of (16)]

periods so that 100 to 1000 photons or more are detected for each position. Under these conditions, a faithful measurement of the photon arrival time is obtained. This approach can easily lead to acquisition times of a complex scene that can be on the order of many seconds or even minutes.

The computational imaging philosophy enables a marked reduction in the number of detected photons needed for 3D imaging (16). In the first-photon imaging approach, only the very first detected photon at each scan location is used, so the acquisition time is limited primarily by the speed of scanning, and any detector dead time coincides with the scanning (17). Using the number of pulses until a photon is detected as an indirect measurement of reflectivity, along with a piecewise-smooth assumption for both reflectivity and depth, a 3D image of a scene is produced after several computational steps, as shown in Fig. 2. This approach builds a strong link between the computational steps and the detailed mechanism of single-photon detection, with various aspects (such as the noise

to very high temporal resolution imaging systems allows reconstruction of a full 3D image of the hidden scene, as conceptually explained in Fig. 3A. A short laser pulse is sent to a scattering surface chosen so as to scatter light behind the obstacle and thus illuminate the hidden scene. The hidden scene will reflect a return echo that will once again hit the first scattering wall and return to the imaging system. An intuitive understanding of the hidden object reconstruction is based on the fact that the locus of points that can give rise to a backscattered signal from a laser spot at a position  $r_l = (x_l, y_l)$  and measured at a given point  $r_i = (x_i, y_i)$  on the wall is given by  $|r - r_l| + |r - r_i| + |z| = c\tau$ . This equation describes an ellipsoid of points that can be recalculated for each detection point on the wall: each of these ellipsoids will overlap only at the points of origin of the (hidden object) scattering. Therefore, by summing over all ellipsoids, one obtains a high “intensity” (proportional to the overlap) in correspondence to the hidden object. With sufficient temporal resolution and additional processing to sharpen the retrieval,



it is possible to reconstruct full 3D shapes: for example, 100 ps is sufficient to resolve centimeter-sized features. Practically, most retrieval techniques aim at iteratively finding an estimate of  $\rho(x, y, z)$ , which represents the physical distribution of the hidden object, from the measured transient image intensity

$$I(x, y, \tau) = \iiint \frac{1}{r_l^2 r_i^2} \delta(|r - r_l| + |r - r_i| + |z| - c\tau) \times \rho(x, y, z) dx dy dz$$

where  $\delta$  represents Dirac's delta function. The first demonstration of this technique was obtained with the use of a streak camera (23) that provides very high 1- to 10-ps temporal resolution but at the expense of relatively long acquisition times (Fig. 3, B and C). Successive demonstrations resorted to single-photon counting to reconstruct 3D images (24) and to perform tasks such as tracking of moving objects (27) and humans, even over very large distances [more than 50 m between the scattering wall and the imaging system (28)]. Recent improvements have demonstrated acquisition times on the order of seconds for a full 3D scene reconstruction by modifying the acquisition scheme. Specifically, photons are collected coaxially [i.e., along the same trajectory as the outgoing laser beam (31)], and, as a result, the measurement integral is simplified to

$$I(x, y, \tau) = \iint \frac{1}{r^4} \delta(r + r_l - c\tau) \rho(x, y, z) dx dy dz$$

where the radiometric factor  $1/r^4$  is now only a function of  $\tau$  and can thus be removed from the integral. Overall, the result of this is that  $I(x, y, \tau)$  reduces to a convolution that substantially decreases the computational retrieval times, paving the way to real-time reconstruction of 3D scenes. This is an example of how imaging hardware and computational techniques have coevolved to create a new imaging capability. It is worth pointing out that recent measurements have shown not only real-time capability but also the capacity for long-distance and full-daylight operation (28, 31), thus moving from proof-of-principle studies to first steps toward deployment in real-world applications in just a few years. An interesting challenge for this field of research starts from the observation that much of the technology involved in NLOS imaging is common with standard, direct line-of-sight of LIDAR (i.e., 3D imaging of environments using laser pulse time-of-flight measurements). In this sense, NLOS imaging has the potential to become a natural extension of LIDAR. In this context, there are clear applications for NLOS imaging, when combined with LIDAR, for urban safety and unmanned vehicles. Additionally, future NASA missions will employ SPAD arrays for LIDAR mapping of planet surfaces, and studies are currently under way to evaluate the potential of NLOS imaging to remotely (e.g., from a satellite) assess the internal structure of underground

caves on planets, with a view toward future human colonization activities (33).

### Enhanced SPAD arrays for imaging in scattering media

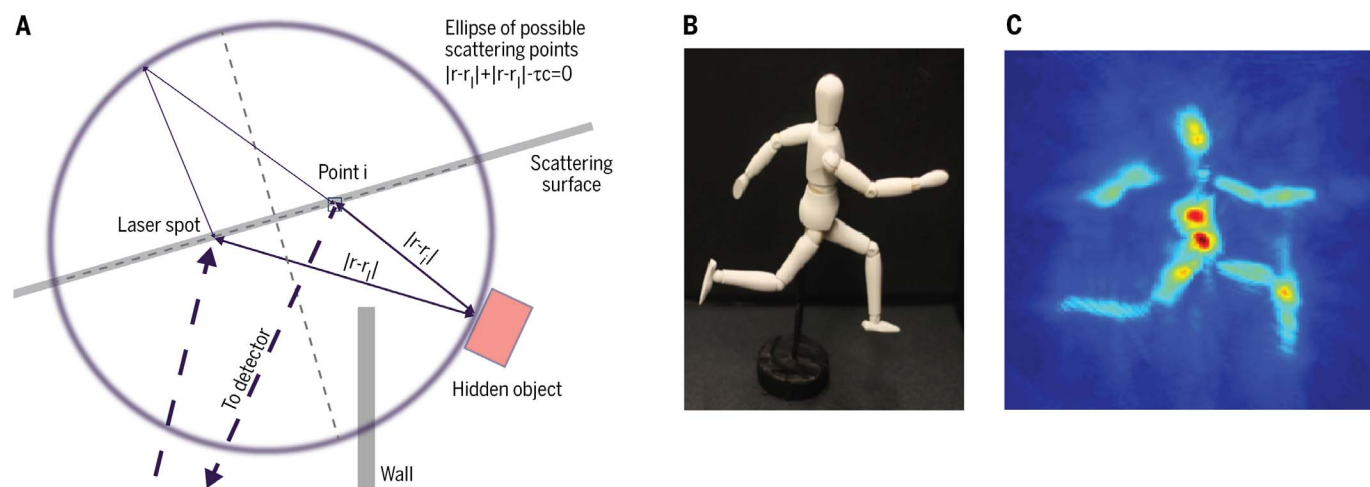
Over the past several years, a number of industrial and academic research groups have been developing a new generation of cameras in which each individual pixel consists of a SPAD. A relatively widespread version of these sensors, often referred to as “quanta imaging sensors,” is operated in binary mode—that is, the sensor pixel generates a “1” when the number of photons is larger than a certain threshold (typically set to just one photon) and generates a “0” otherwise (34–38). Each frame therefore has a single-bit depth. To build an image, multiple frames must be added together. This operation mode brings some advantages: Aside from the single-photon sensitivity, one can add as many frames as desired so as to achieve very high bit depths (dynamic ranges) that are not attainable with standard complementary metal-oxide semiconductor cameras. Moreover, the single-bit nature of the acquisition permits very high frame acquisition rates (rates up to 100 kHz have been reported) (39).

Progress has also been made in the full on-chip integration of TCSPC electronics, thus providing the additional functionality of temporal resolutions as low as 50 ps (21, 40–43). This implies that, when combined with a high repetition rate laser for the active illumination of the scene, video rates reaching up to 20 billion frames per second can be achieved (44). This remarkable performance can be better appreciated when expressed in terms of the actual imaging capability. At such frame rates, light propagates just 1.5 cm between successive frames, which implies that it is possible to actually freeze light in motion in much the same way that standard high-speed cameras can freeze the motion of a supersonic bullet. The first images of light in flight were shown in the late 1960s via nonlinear optical gating methods (45–47), but the first camera-based measurements were only recently demonstrated with the use of a streak camera (48). More recent measurements based on SPAD arrays have allowed the first capture of light pulses propagating in free space with total acquisitions times on the order of seconds or less (44). SPAD array cameras have also been used to directly image laser pulse propagation through optical fibers: Beyond their direct applications (e.g., estimation of physical parameters of optical fibers), these measurements combined a fusion of single-photon data with hyperspectral imaging over several different wavelengths (discussed below) and computational processing through which the 32-pixel-by-32-pixel resolution was successfully up-sampled by using the temporal axis to re-interpolate the pulse trajectory in the  $(x, y)$  spatial plane (49).

The ability to capture simultaneously spatial and high-resolution temporal information at very low light levels with SPAD cameras has recently been applied to other difficult imaging problems, such as imaging and sensing through scattering and turbid media. For example, Pavia *et al.* have

applied inverse retrieval methods in combination with spatial and temporal information from a linear SPAD array for tomographic reconstruction of objects hidden in murky water (50). More recently, Heshmat and co-workers have acquired data with a planar SPAD array and reconstructed various shapes of objects obscured by a thick tissue phantom (51). Their technique was referred to as “All Photons Imaging,” directly underlining the importance of the photon time-of-flight information that is recorded by the single-photon camera. We note that such approaches do not explicitly account for the physical origins of the data. For example, temporal and spatial information are placed on equal footing and enter in the retrieval process without incorporation of statistical models for timing jitter or surface reflectivity of the objects. Future SPAD imaging will benefit from incorporating accurate spatiotemporal statistical models for sources, photon transport medium, and photon detectors. In the broad regime of strong scattering, the camera will typically record an indistinct, diffuse illumination transmitted through the medium or reflected from the scene with little or no obvious information about any objects hidden behind or inside the scattering medium. Computational techniques are thus required to actually retrieve details about the object. This field of research is of particular interest for a number of applications such as medical diagnostics and imaging, as well as sensing and imaging through fog.

With the emergence of imaging systems for autonomous underwater vehicles, unmanned aerial vehicles, robots, or cars, rain and fog present important challenges that must be overcome. Sonar is a well-established technology for long range underwater imaging, but it can suffer from low spatial resolution limited by the physics of sound propagation in the medium. Although high-power optical solutions can be used for short-range imaging in relatively clear water, the presence of underwater scatterers between the active imaging system and the scene (e.g., the seabed) usually produce large quantities of reflected photons that can mask the returns from the scene of interest. By using a pulsed illumination source combined with sensitive single-photon detectors, it is possible to discriminate the photons reflected because of scattering in the water from those (an extremely small refraction) that actually reach the surfaces of interest. For instance, Maccarone *et al.* have demonstrated the ability to image underwater up to eight attenuation lengths (52). When combining this cutting-edge technology with advanced statistical methods inspired by our previous work (53), substantial performance improvements could be achieved in terms of 3D reconstruction and estimation of the surface reflectivity by accounting for the distance-induced signal loss (54). On a related note, efforts for 3D reconstruction of terrestrial scenes at long distances suffer from limitations similar to those described above. Even if the measurements are performed under favorable (e.g., dry) conditions, the recorded signals can be considerably affected by atmospheric



**Fig. 3. Non-line-of-sight imaging.** (A) Basic geometry of the problem. A laser illuminates a scattering surface and scatters light around a wall that hides an object from the direct line of sight. The return signal backscattered from the hidden object is detected at a point “i” on the scattering surface. This geometry, with a single observation point, defines

an ellipsoid of possible locations for the object. Detection of the time-resolved transient images at multiple points on the surface allows reconstruction of the location or even the full 3D shape of the object. (B) An example of a hidden object with its reconstruction shown in (C). [Panels (B) and (C) adapted from (21)]

turbulence (55–57) and solar illumination (58, 59). Again, marked improvements in detection accuracy (60) and maximal observable range (61) have been obtained via the use of adapted computational tools. The problem becomes even more acute in the presence of fog, which is a major concern for the next generation of automated cars. It has been demonstrated that it is technically possible to detect and analyze fog patches over long distances, provided that the laser power is sufficient to ensure a nonzero probability of photon reflection and a long enough acquisition time (62, 63). In the automotive context, where the acquisition time is intrinsically limited by the vehicle displacement velocity, more robust and computationally efficient strategies have been recently proposed (51, 64), and it is clear that future imaging systems will incorporate computational models for both the propagation physics of the medium and physical properties of the detector.

### Multispectral single-photon imaging

Multispectral and hyperspectral imaging, which are extensions of classical color (RGB) imaging, consist of imaging a scene using multiple wavelengths (from four to several hundreds or even thousands in hyperspectral images). These modalities have benefited from a robust body of research spanning more than 35 years, from the data collection community (65–67), and, more importantly, from the data processing and analysis community (68–73). Indeed, such modalities can be associated with a wide variety of computational problems, ranging from image acquisition (compressive sampling), restoration (denoising/deblurring, superresolution), segmentation (classification) to source separation (spectral unmixing), object/anomaly detection, and data fusion (e.g., pansharpening). Though the main applica-

tions using (passive) multispectral imaging are in Earth and space observation, the proven benefits of imaging with multiple wavelengths simultaneously have enabled its application in the food industry (66, 74) and a broader range of applications such as diagnostic dermatology (75, 76). Active multispectral imaging is less sensitive to ambient illumination than passive imaging, which requires data acquisition under daylight condition (e.g., for Earth observation). Without fast timing capabilities, however, multi- and hyperspectral imagers are only able to provide 2D intensity profiles and are thus poorly adapted to analysis of multilayered 3D structures such as forest canopies. Multispectral LIDAR is a promising modality that allows for joint extraction of geometric (as single-wavelength LIDAR) and spectral (as passive multispectral images) information from the scene while avoiding data registration issues potentially induced by the fusion of heterogeneous sensors. Wallace *et al.* have demonstrated that it is possible to use multispectral single-photon LIDAR (MSPL) to remotely infer the spatial composition (leaves and branches) and the health of trees using only four wavelengths (77). More recently, new experiments have been designed to image up to 33 wavelengths (500 to 820 nm) in free space (78) and 16 wavelengths underwater (79). As a consequence, we have witnessed the development of algorithms inspired from passive hyperspectral imagery (3D datacubes) for analysis of MSPL data (4D datacubes).

For instance, Bayesian methods have been proposed to cluster, in an unsupervised manner, spectrally similar objects while estimating their range from photon-starved MSPL data (80). This work was further developed (81, 82) to classify pixels on the basis of their spectral profiles in photon-starved regimes down to one photon per

pixel and per spectral band, on average (Fig. 4). Such results are possible only by efficiently combining a highly sensitive raster-scanning single-photon system that allows for submillimeter range resolution with hierarchical Bayesian models able to capture the intrinsic, yet faint, structures (e.g., spatial and spectral correlations) of the data. A notable improvement has been demonstrated by using simulation methods (see next section) to reconstruct scenes (range and reflectivity profiles) with as few as four photons per pixel (with four spectral bands and one photon per pixel, on average) (81).

Spectral unmixing presents another challenging problem in the use of multi- and hyperspectral data for identification and quantification of materials or components present in the observed scene. Spectral unmixing can lead to improved classification by accounting for the fact that several mixed materials can be observed in a given pixel. Spectral unmixing methods allow for sub-pixel material quantification, which is particularly important for long-range imaging scenarios in which the divergence of the laser beam cannot be neglected. We developed a computational method for quantifying and locating 15 known materials from MSPL data consisting of 33 spectral bands while detecting additional (potentially unknown) materials present in the scene (78). Again, this work demonstrated the possibility of material quantification and anomaly detection with as little as 1 photon per pixel and per spectral band, on average. It also illustrated how Bayesian modeling can be used for uncertainty quantification—e.g., for providing confidence intervals associated with estimated range profiles. As mentioned above, although the most recent single-photon detectors are very attractive because of their high temporal resolution, their application to information extraction from wide-area scenes is hindered



by long acquisition times associated with raster scanning strategies. This is particularly limiting when several wavelengths are acquired in a sequential manner. To address this problem, compressive sampling strategies have been investigated to achieve faster MSPL data acquisition (83, 84). Although computational methods for image scanning systems have been proposed, whereby a random number of spectral bands can be probed in a given pixel, the most promising results have been obtained with a simulated mosaic filter (four wavelengths) whose implementation within a SPAD array could allow for the simultaneous acquisition of multiple pixels and fast reconstruction of range and reflectivity profiles. These results show how advanced computational methods can be used to enhance information extraction from imaging systems and also improve the design of future detectors and detector arrays.

### Computational methods in the photon-starved regime

From a mathematical perspective, computational imaging is formulated as finding a mapping that reconstructs a set of parameters  $\mathbf{x}$ , which may have a physical meaning (or not), from a set of measurements  $\mathbf{y}$  recorded by an imaging or sensing device. These parameters can take continuous values (e.g., light field intensities, object positions, and velocities) or discrete values (e.g., the number objects, binary values representative of the presence or absence of objects). Two main families of methods can be adopted to design algorithms for data analysis—namely, supervised machine learning approaches and statistical signal processing approaches—although hybrid methods can also be used. The choice of the most suitable approaches depends primarily on the complexity of the computational model, as well as the computational budget available (i.e., the

expected processing time, data storage limitations, and the desired quality of the information extracted). Supervised machine learning (including deep learning) approaches are particularly well suited for applications where a sufficient quantity of ground truth data or reference data is available (85–88). Such methods rely on a two-stage process, which consists of the training stage and the test stage. Starting from a forward model  $\mathbf{y} \approx g(\mathbf{x})$ , relating the measured data  $\mathbf{y}$  to the unknown source parameters  $\mathbf{x}$ , the training stage uses a set of measurements and corresponding parameters to learn the inverse mapping  $h(\cdot)$  between the measurements and the set of parameters to be recovered—i.e., it fits an inverse model  $\mathbf{x} \approx h(\mathbf{y})$ . In contrast to model-based statistical methods, data-driven machine learning approaches do not rely on the knowledge a forward model  $g(\cdot)$ . Thus, these methods can often be applied to complex problems where the forward model is unknown or too complicated to be derived analytically but for which plenty of training data are available. The training stage controls the quality of the estimation of the mapping  $h(\cdot)$ , which in turn depends on the representational power of the machine learning algorithm and on the quality and diversity of the training samples. Machine learning approaches have been successfully applied to imaging applications such as imaging through multimode fibers (85), lensless imaging of phase objects (86), and identification of a human pose from behind a diffusive screen (87). SPAD cameras have been specifically applied to identifying both the positions and identities of people hidden behind a wall (88), imaging through partially transmissive (89) and scattering media (51), and profiling camouflaged targets (90). The design of reliable machine learning approaches is currently limited by high generalization error (i.e., machines must be retrained for different acquisition scenarios)

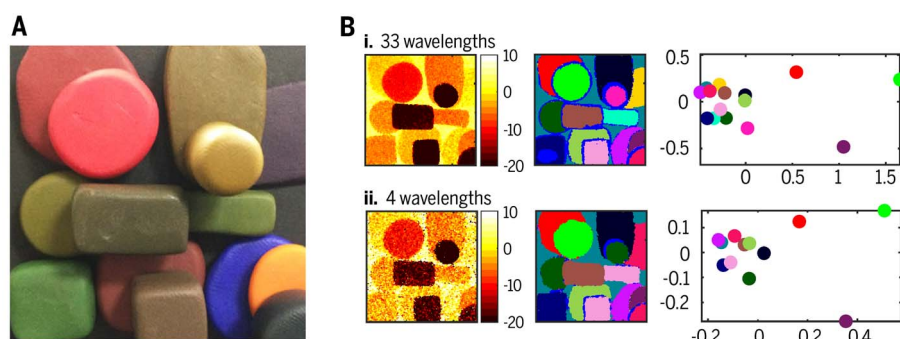
and a lack of ground truth information about the sources or the medium.

Statistical model-based methods can thus be more attractive than data-driven machine learning approaches for photon-limited imaging, as a mathematical forward model  $g(\cdot)$  can be combined with a statistical noise model to better fit the data. Physical considerations, such as light transport theory through the medium and the detector, can often guide model choice, although non-physically inspired approximations can be used to make the model fitting algorithm more computationally tractable. When there is measurement uncertainty and noise, the forward model can be better characterized by the conditional probability distribution  $f(\mathbf{y}|\mathbf{x})$ , which describes the statistical variation of the measurements  $\mathbf{y}$  for a given source parameter value  $\mathbf{x}$ . For fixed value  $\mathbf{y}$ , the function  $l_y(\mathbf{x}) = f(\mathbf{y}|\mathbf{x})$ , called the likelihood function, quantifies the likelihood that the source value  $\mathbf{x}$  generated the observed value  $\mathbf{y}$ . The maximum likelihood principle forms an estimate of  $\mathbf{x}$  from  $\mathbf{y}$  by maximizing the likelihood over  $\mathbf{x}$ . However, the maximum likelihood estimate is often not unique in high-dimensional inverse problems such as imaging. Fortunately, additional information about  $\mathbf{x}$  (e.g., a priori knowledge about positivity, smoothness, or sparsity) is often available and can be used to improve on the maximum likelihood estimate. For instance, suppose  $\phi$  is a regularization function such that  $\phi(\mathbf{x})$  is small when  $\mathbf{x}$  complies with a priori knowledge and is large otherwise. Then it is possible to recover  $\mathbf{x}$  by minimizing the cost function  $C_y(\mathbf{x}) = -\log(l_y(\mathbf{x})) + \phi(\mathbf{x})$ . If  $\phi(\mathbf{x})$  can be associated with a proper density  $f(\mathbf{x}) \propto e^{-\phi(\mathbf{x})}$ , called the prior distribution, this penalized likelihood estimation strategy can be interpreted in the Bayesian formalism as a maximum a posteriori estimation procedure. In other words, the above minimization is equivalent to maximizing the posterior density of  $\mathbf{x}$

$$f(\mathbf{x}|\mathbf{y}) = f(\mathbf{y}|\mathbf{x})f(\mathbf{x})/f(\mathbf{y})$$

where  $f(\mathbf{y})$  is a density that does not depend on  $\mathbf{x}$ . These and related likelihood-based approaches have been adopted by many researchers studying low photon imaging (17, 18, 83).

The Bayesian formalism allows for the observed data to be combined with additional information in a principled manner. This also allows so-called a posteriori measures of uncertainty to be derived. However, such measures cannot be computed analytically in most practical applications because of the complexity of accurate spatial correlation models, and they often must be approximated using high-dimensional integration. A considerable advantage may be gained from computationally simple pixel-by-pixel adaptation (91), and a classical approach thus consists of approximating these measures (e.g., a posteriori variances/covariances or confidence intervals) using variational approximations or simulation techniques. Markov chain Monte



**Fig. 4. Computational inverse probability methods to spectrally classify and depth-resolve objects in a scene from photon-starved multispectral LIDAR images.** The scene (A) was composed of 14 clay materials of different colors. The recorded images consist of a 200-pixel-by-200-pixel area (scanned target areas were approximately 50 mm by 50 mm), and the targets were placed 1.85 m from the system. In (B), the first column depicts the estimated depth profile (in millimeters), the reference range being arbitrarily set to the backplane range. The second column shows color classification maps, and the third column depicts the spectral signatures of the most prominent classes, projected onto the first and second axes obtained by using principal components analysis. Each of these subplots illustrates the similarity between the estimated spectral signatures. Rows (i) and (ii) in (B) depict results obtained with an average of one detected photon per pixel, for each spectral band, with 33 and 4 bands, respectively. [Adapted with permission from (73)]

Carlo methods are particularly well adapted for inference in difficult scenarios for which the cost function or posterior distribution of interest has multiple modes and multiple solutions potentially admissible. For instance, such methods have been successfully applied to object detection (60), and joint material identification and anomaly detection (78) from low-flux single-photon LIDAR measurements.

## Conclusions

Considering the rapid advance in imaging cameras and sensors together with a leap forward in computational capacity, we see enormous potential for innovation over the next several years. The main challenge—or the main opportunity—at this stage is the codevelopment of sensors and computational algorithms built around the physical processes of the photon transport and detection mechanisms. We have provided examples showing progress in this direction, ranging from first-photon imaging techniques to photon-starved hyperspectral imaging. The trend seen in commercial cameras between 2000 and 2015, characterized by a constant drive toward higher pixel counts, has slowly subsided, giving way to a different approach whereby both performance and functionality are increased by combining multiple sensors through computational processing. Obvious examples are recent advances in cell-phone technology, arguably one of the main drivers behind imaging technology, that now boasts multiple cameras and lenses providing depth perception, improved signal-to-noise ratio, and other functionalities such as 3D face recognition. With SPAD cameras also gradually making their appearance on the commercial scene, single-photon imaging and computational techniques offer a promising avenue for future innovation in situations where previously imaging was thought to be impossible. We have briefly discussed examples such as imaging through denser scattering media (the human body or fog) and full 3D imaging of scenes around a corner or beyond the direct line of sight. These examples are particularly relevant in demonstrating the progress that can be made when the photon transport models and computational approaches are integrated with the new generation of photon detectors. The expectation is that over the next several years we will witness substantial growth of computational imaging methods, driven by and also driving new technologies such as single-photon SPAD arrays that will revolutionize nearly every aspect of human activity, ranging from medical diagnostics to urban safety and space missions.

## REFERENCES AND NOTES

1. T. B. Pittman, Y. H. Shih, D. V. Strekalov, A. V. Sergienko, Optical imaging by means of two-photon quantum entanglement. *Phys. Rev. A* **52**, R3429–R3432 (1995). doi: [10.1103/PhysRevA.52.R3429](#); pmid: [9912767](#)
2. J. H. Shapiro, R. W. Boyd, The physics of ghost imaging. *Quantum Inform. Process.* **11**, 949–993 (2012). doi: [10.1007/s1128-011-0356-5](#)
3. J. Shapiro, Computational ghost imaging. *Phys. Rev. A* **78**, 061802 (2008). doi: [10.1103/PhysRevA.78.061802](#)
4. M. F. Duarte et al., Single-pixel imaging via compressive sampling. *IEEE Signal Process. Mag.* **25**, 83–91 (2008). doi: [10.1109/MSP.2007.914730](#)
5. D. Shin, J. H. Shapiro, V. K. Goyal, Performance analysis of low-flux least-squares single-pixel imaging. *IEEE Signal Process. Lett.* **23**, 1756–1760 (2016). doi: [10.1109/LSP.2016.2617329](#)
6. P. A. Morris, R. S. Aspden, J. E. Bell, R. W. Boyd, M. J. Padgett, Imaging with a small number of photons. *Nat. Commun.* **6**, 5913 (2015). doi: [10.1038/ncomms5913](#); pmid: [25557090](#)
7. A. Gatti, E. Brambilla, M. Bache, L. A. Lugiato, Correlated imaging, quantum and classical. *Phys. Rev. A* **70**, 013802 (2004). doi: [10.1103/PhysRevA.70.013802](#)
8. M. P. Edgar et al., Simultaneous real-time visible and infrared video with single-pixel detectors. *Sci. Rep.* **5**, 10669 (2015). doi: [10.1038/srep10669](#); pmid: [26001092](#)
9. G. Satat, M. Tancik, R. Raskar, Lensless imaging with compressive ultrafast sensing. *IEEE Trans. Computat. Imaging* **3**, 398–407 (2017). doi: [10.1109/TCI.2017.2684624](#)
10. G. A. Howland, D. J. Lum, M. R. Ware, J. C. Howell, Photon counting compressive depth mapping. *Opt. Express* **21**, 23822–23837 (2013). doi: [10.1364/OE.21.023822](#); pmid: [24104293](#)
11. M.-J. Sun et al., Single-pixel three-dimensional imaging with time-based depth resolution. *Nat. Commun.* **7**, 12010–12016 (2016). doi: [10.1038/ncomms12010](#); pmid: [27377197](#)
12. D. B. Phillips et al., Adaptive foveated single-pixel imaging with dynamic supersampling. *Sci. Adv.* **3**, e1601782 (2017). doi: [10.1126/sciadv.1601782](#); pmid: [28439538](#)
13. D. V. O'Connor, D. Phillips, *Time-Correlated Single Photon Counting* (Academic Press, 1984).
14. W. Becker, *Advanced Time-Correlated Single Photon Counting Techniques* (Springer Series in Chemical Physics, Springer, 2005).
15. S. Pellegrini, G. S. Buller, J. M. Smith, A. M. Wallace, S. Cova, Laser-based distance measurement using picosecond resolution time-correlated single-photon counting. *Meas. Sci. Technol.* **11**, 712–716 (2000). doi: [10.1088/0957-0233/11/6/314](#)
16. D. B. Lindell, M. O'Toole, G. Wetzstein, Single-photon 3D imaging with deep sensor fusion. *ACM Trans. Graph.* **37**, 113 (2018).
17. A. Kirmani et al., First-photon imaging. *Science* **343**, 58–61 (2014). doi: [10.1126/science.1246775](#); pmid: [24292628](#)
18. D. Shin, A. Kirmani, V. K. Goyal, J. H. Shapiro, Photon-efficient computational 3D and reflectivity imaging with single-photon detectors. *IEEE Trans. Computat. Imaging* **1**, 112–125 (2015). doi: [10.1109/TCI.2015.2453093](#)
19. J. Rapp, V. K. Goyal, A few photons among many: Unmixing signal and noise for photon-efficient active imaging. *IEEE Trans. Computat. Imaging* **3**, 445–459 (2017). doi: [10.1109/TCI.2017.2706028](#)
20. C. F. Higham, R. Murray-Smith, M. J. Padgett, M. P. Edgar, Deep learning for real-time single-pixel video. *Sci. Rep.* **8**, 2369 (2018). doi: [10.1038/s41598-018-20521-y](#); pmid: [29403059](#)
21. F. Villa et al., CMOS imager with 1024 SPADs and TDCs for single-photon timing and 3-D time-of-flight. *IEEE J. Sel. Top. Quantum Electron.* **20**, 364–373 (2014). doi: [10.1109/JSTQE.2014.2342197](#)
22. D. Shin et al., Photon-efficient imaging with a single-photon camera. *Nat. Commun.* **7**, 12046 (2016). doi: [10.1038/ncomms12046](#); pmid: [27338821](#)
23. A. Velten et al., Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nat. Commun.* **3**, 745 (2012). doi: [10.1038/ncomms1747](#); pmid: [22434188](#)
24. M. Buttafava, J. Zeman, A. Tosi, K. Eliceiri, A. Velten, Non-line-of-sight imaging using a time-gated single photon avalanche diode. *Opt. Express* **23**, 20997–21011 (2015). doi: [10.1364/OE.23.020997](#); pmid: [26367952](#)
25. M. Laurenzis, A. Velten, Non line-of-sight laser gated viewing of scattered photons. *Opt. Eng.* **53**, 023102 (2014). doi: [10.1117/1.OE.53.2.023102](#)
26. J. Klein, C. Peters, J. Martin, M. Laurenzis, M. B. Hullin, Tracking objects outside the line of sight using 2D intensity images. *Sci. Rep.* **6**, 32491 (2016). doi: [10.1038/srep32491](#); pmid: [27577969](#)
27. G. Gariépy, F. Tonolini, R. Henderson, J. Leach, D. Faccio, Detection and tracking of moving objects hidden from view. *Nat. Photonics* **10**, 23–26 (2016). doi: [10.1038/nphoton.2015.234](#)
28. S. Chan, R. E. Warburton, G. Gariépy, J. Leach, D. Faccio, Non-line-of-sight tracking of people at long range. *Opt. Express* **25**, 10109–10117 (2017). doi: [10.1364/OE.25.010109](#); pmid: [28468386](#)
29. N. Naik, S. Zhao, A. Velten, R. Raskar, K. Bala, Single view reflectance capture using multiplexed scattering and time-of-flight imaging. *ACM Trans. Graph.* **30**, 171 (2011). doi: [10.1145/2070781.2024205](#)
30. R. Pandharkar, A. Velten, A. Bardagj, E. Lawson, M. Bawendi, R. Raskar, "Estimating motion and size of moving non-line-of-sight objects in cluttered environments" in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2011)* (IEEE, 2011), pp. 265–272.
31. M. O'Toole, D. B. Lindell, G. Wetzstein, Confocal non-line-of-sight imaging based on the light-cone transform. *Nature* **555**, 338–341 (2018). doi: [10.1038/nature25489](#); pmid: [29513650](#)
32. A. Kirmani, H. Jeelani, V. Montazerhodjat, V. K. Goyal, Diffuse imaging: Creating optical images with unfocused time-resolved illumination and sensing. *IEEE Signal Process. Lett.* **19**, 31–34 (2012). doi: [10.1109/LSP.2011.2174222](#)
33. P. Boston et al., "Human utilization of subsurface extraterrestrial environments: Final report" (NASA Institute for Advanced Concepts, 2004); [www.niac.usra.edu/files/studies/final\\_report/710Boston.pdf](#)
34. R. Fossum, "What to do with sub-diffraction-limit (SDL) pixels? A proposal for a gigapixel digital film sensor (DFS)" in *2005 IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors* (International Image Sensor Society, 2005), pp. 214–217.
35. S. Masoodian, A. Rao, J. Ma, K. Odame, E. R. Fossum, A 2.5 pJ/binary image sensor as a pathfinder for quanta image sensors. *IEEE Trans. Electron Dev.* **63**, 100–105 (2016). doi: [10.1109/TED.2015.2457418](#)
36. Y. Altmann, R. Aspden, M. Padgett, S. McLaughlin, A Bayesian approach to denoising of single-photon binary images. *IEEE Trans. Computational Imaging* **3**, 460–471 (2017). doi: [10.1109/TCI.2017.2703900](#)
37. J. H. Choi, O. A. Elgendy, S. H. Chan, "Image reconstruction for Quanta Image Sensors using deep neural networks" in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, (IEEE, 2018), pp. 6543–6547.
38. O. A. Elgendy, S. H. Chan, Optimal threshold design for Quanta image sensors. *IEEE Trans. Computat. Imaging* **4**, 99–111 (2018). doi: [10.1109/TCI.2017.2781185](#)
39. M. Antolovic, S. Burri, C. Bruschini, R. Hoebe, E. Charbon, Nonuniformity analysis of a 65k pixel CMOS SPAD imager. *IEEE Trans. Electron Dev.* **63**, 57–64 (2016). doi: [10.1109/TED.2015.2458295](#)
40. C. Veerappan, J. Richardson, R. Walker, D.-U. Li, M. W. Fishburn, Y. Maruyama, D. Stoppa, F. Borghetti, M. Gersback, R. K. Henderson, E. Charbon, A 160 × 128 single-photon image sensor with on-pixel 55 ps 10 bit time-to-digital converter. *IEEE International Solid-State Circuits Conference (ISSCC) Digest of Technical Papers* (IEEE, 2011), pp. 312–314.
41. I. Vorricu, R. Carmona-Galan, A. Rodriguez-Vazquez, Arrayable voltage-controlled ring-oscillator for direct time-of-flight image sensors. *IEEE Trans. Circuits Syst. I Regul. Pap.* **64**, 2821–2834 (2017). doi: [10.1109/TCSI.2017.2706324](#)
42. L. Gasparini, M. Zarghami, H. Xu, L. Parmesan, M. M. Garcia, M. Unterhahr, B. Bessire, A. Stefanov, D. Stoppa, M. Perenzoni, "A 32x32-pixel time-resolved single-photon image sensor with 44.6 μm pitch and 19.48% fill-factor with on-chip row/frame skipping features reaching 800kHz observation rate for quantum physics applications" in *IEEE ISSCC* (IEEE, 2018), pp. 98–100.
43. D. B. Lindell, M. O'Toole, G. Wetzstein, "Towards transient imaging at interactive rates with single-photon detectors," in *IEEE International Conference on Computational Photography (ICCP)* (IEEE, 2018), pp. 1–8.
44. G. Gariépy et al., Erratum: Single-photon sensitive light-in-flight imaging. *Nat. Commun.* **6**, 6408 (2015). doi: [10.1038/ncomms7021](#); pmid: [25711544](#)
45. J. A. Giordmaine, P. M. Rentzepis, S. L. Shapiro, K. W. Wecht, Two photon excitation of fluorescence by picosecond light pulses. *Appl. Phys. Lett.* **11**, 216–218 (1967). doi: [10.1063/1.1755105](#)
46. M. A. Duguay, J. W. Hansen, An ultrafast light gate. *Appl. Phys. Lett.* **15**, 192–194 (1969). doi: [10.1063/1.1652962](#)
47. M. A. Duguay, A. T. Mattick, Ultrahigh speed photography of picosecond light pulses and echoes. *Appl. Opt.* **10**, 2162–2170 (1971). pmid: [20111288](#)
48. A. Velten et al., Femto-photography: Capturing and visualizing the propagation of light. *ACM Trans. Graph.* **32**, 1–8 (2013). doi: [10.1145/2461912.2461928](#)
49. R. Warburton et al., Observation of laser pulse propagation in optical fibers with a SPAD camera. *Sci. Rep.* **7**, 43302 (2017). doi: [10.1038/srep43302](#); pmid: [28266554](#)



50. J. M. Pavia, M. Wolf, E. Charbon, Single-photon avalanche diode images applied to near-infrared imaging. *IEEE J. Sel. Top. Quantum Electron.* **20**, 3800908 (2014).
51. G. Satat, B. Heshmat, D. Raviv, R. Raskar, All photons imaging through volumetric scattering. *Sci. Rep.* **6**, 33946 (2016). doi: [10.1038/srep33946](https://doi.org/10.1038/srep33946); pmid: [27683065](https://pubmed.ncbi.nlm.nih.gov/27683065/)
52. A. Maccarone *et al.*, Underwater depth imaging using time-correlated single-photon counting. *Opt. Express* **23**, 33911–33926 (2015). doi: [10.1364/OE.23.033911](https://doi.org/10.1364/OE.23.033911); pmid: [26832050](https://pubmed.ncbi.nlm.nih.gov/26832050/)
53. Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, S. McLaughlin, Lidar waveform-based analysis of depth images constructed using sparse single photon data. *IEEE Trans. Image Process.* **25**, 1935–1946 (2016). doi: [10.1109/TIP.2016.2526784](https://doi.org/10.1109/TIP.2016.2526784); pmid: [26866984](https://pubmed.ncbi.nlm.nih.gov/26866984/)
54. A. Halimi, A. Maccarone, A. McCarthy, S. McLaughlin, G. S. Buller, Object depth profile and reflectivity restoration from sparse single-photon data acquired in underwater environments. *IEEE Trans. Computat. Imaging* **3**, 472–484 (2017). doi: [10.1109/TCI.2017.2669867](https://doi.org/10.1109/TCI.2017.2669867)
55. P. W. Milonni, J. H. Carter, C. G. Peterson, R. J. Hughes, Effects of propagation through atmospheric turbulence on photon statistics. *J. Opt. B Quantum Semiclassical Opt.* **6**, S742 (2004). doi: [10.1088/1464-4266/6/8/018](https://doi.org/10.1088/1464-4266/6/8/018)
56. I. Capraro *et al.*, Impact of turbulence in long range quantum and classical communications. *Phys. Rev. Lett.* **109**, 200502 (2012). doi: [10.1103/PhysRevLett.109.200502](https://doi.org/10.1103/PhysRevLett.109.200502); pmid: [23215467](https://pubmed.ncbi.nlm.nih.gov/23215467/)
57. M. Henriksson, L. Sjöqvist, Scintillation index measurement using time-correlated single-photon counting laser radar. *Opt. Eng.* **53**, 081902 (2014). doi: [10.1117/1.OE.53.8.081902](https://doi.org/10.1117/1.OE.53.8.081902)
58. A. McCarthy *et al.*, Kilometer-range depth imaging at 1550 nm wavelength using an InGaAs/InP single-photon avalanche diode detector. *Opt. Express* **21**, 22098–22113 (2013). doi: [10.1364/OE.21.022098](https://doi.org/10.1364/OE.21.022098); pmid: [24104102](https://pubmed.ncbi.nlm.nih.gov/24104102/)
59. A. McCarthy *et al.*, Kilometer-range, high resolution depth imaging via 1560 nm wavelength single-photon detection. *Opt. Express* **21**, 8904–8915 (2013). doi: [10.1364/OE.21.008904](https://doi.org/10.1364/OE.21.008904); pmid: [23571981](https://pubmed.ncbi.nlm.nih.gov/23571981/)
60. Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, S. McLaughlin, Robust Bayesian target detection algorithm for depth imaging from sparse single-photon data. *IEEE Trans. Computat. Imaging* **2**, 456–467 (2016).
61. A. M. Pawlikowska, A. Halimi, R. A. Lamb, G. S. Buller, Single-photon three-dimensional imaging at up to 10 kilometers range. *Opt. Express* **25**, 11919–11931 (2017). doi: [10.1364/OE.25.011919](https://doi.org/10.1364/OE.25.011919); pmid: [28788749](https://pubmed.ncbi.nlm.nih.gov/28788749/)
62. J. Zhu *et al.*, Demonstration of measuring sea fog with an SNSPD-based Lidar system. *Sci. Rep.* **7**, 15113 (2017). doi: [10.1038/s41598-017-15429-y](https://doi.org/10.1038/s41598-017-15429-y); pmid: [29118415](https://pubmed.ncbi.nlm.nih.gov/29118415/)
63. B. Du *et al.*, High-speed photon-counting laser ranging for broad range of distances. *Sci. Rep.* **8**, 4198 (2018). doi: [10.1038/s41598-018-22675-1](https://doi.org/10.1038/s41598-018-22675-1); pmid: [29520022](https://pubmed.ncbi.nlm.nih.gov/29520022/)
64. G. Satat, M. Tancik, R. Raskar, “Towards photography through realistic fog” in *IEEE ICCP* (IEEE, 2018), pp. 1–10.
65. F. H. Goetz, Three decades of hyperspectral remote sensing of the Earth: A personal view. *Remote Sens. Environ.* **113**, S5–S16 (2009). doi: [10.1016/j.rse.2007.12.014](https://doi.org/10.1016/j.rse.2007.12.014)
66. A. A. Gowen, C. P. O'Donnell, P. J. Cullen, G. Downey, J. M. Frias, Hyperspectral imaging - an emerging process analytical tool for food quality and safety control. *Trends Food Sci. Technol.* **18**, 590–598 (2007). doi: [10.1016/j.tifs.2007.06.001](https://doi.org/10.1016/j.tifs.2007.06.001)
67. G. P. Asner *et al.*, Carnegie airborne observatory: In-flight fusion of hyperspectral imaging and waveform light detection and ranging for three-dimensional studies of ecosystems. *J. Appl. Remote Sens.* **1**, 013536 (2007). doi: [10.1117/1.2794018](https://doi.org/10.1117/1.2794018)
68. D. Landgrebe, Hyperspectral image data analysis. *IEEE Signal Process. Mag.* **19**, 17–28 (2002). doi: [10.1109/79.974718](https://doi.org/10.1109/79.974718)
69. A. Plaza *et al.*, Recent advances in techniques for hyperspectral image processing. *Remote Sens. Environ.* **113**, S110–S122 (2009). doi: [10.1016/j.rse.2007.07.028](https://doi.org/10.1016/j.rse.2007.07.028)
70. J. M. Bioucas-Dias *et al.*, Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **5**, 354–379 (2012). doi: [10.1109/JSTARS.2012.2194696](https://doi.org/10.1109/JSTARS.2012.2194696)
71. M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, J. C. Tilton, Advances in spectral-spatial classification of hyperspectral images. *Proc. IEEE* **101**, 652–675 (2013). doi: [10.1109/JPROC.2012.2197589](https://doi.org/10.1109/JPROC.2012.2197589)
72. N. Dobigeon *et al.*, Nonlinear unmixing of hyperspectral Images: Models and algorithms. *IEEE Signal Process. Mag.* **31**, 82–94 (2014). doi: [10.1109/MSP.2013.2279274](https://doi.org/10.1109/MSP.2013.2279274)
73. L. Loncan *et al.*, Hyperspectral pansharpening: A review. *IEEE Geosc. and Rem. Sens. Mag.* **3**, 27–46 (2015). doi: [10.1109/MGRS.2015.2440094](https://doi.org/10.1109/MGRS.2015.2440094)
74. D.-W. Sun, *Hyperspectral Imaging for Food Quality Analysis and Control* (Academic Press, 2010)
75. R. Koprowski, P. Olczyk, Segmentation in dermatological hyperspectral images: Dedicated methods. *Biomed. Eng. Online* **15**, 97 (2016). doi: [10.1186/s12938-016-0219-5](https://doi.org/10.1186/s12938-016-0219-5); pmid: [27535027](https://pubmed.ncbi.nlm.nih.gov/27535027/)
76. F. Vasefi, N. MacKinnon, D. L. Farkas, “Chapter 16 - Hyperspectral and Multispectral Imaging in Dermatology” in *Imaging in Dermatology*, M. R. Hamblin, P. Avci, G. K. Gupta, Eds. (Academic Press, 2016), pp. 187–201.
77. A. M. Wallace *et al.*, Design and Evaluation of Multispectral Lidar for the recovery of arboreal parameters. *IEEE Trans. Geosci. Remote Sens.* **52**, 4942–4954 (2014). doi: [10.1109/TGRS.2013.2285942](https://doi.org/10.1109/TGRS.2013.2285942)
78. Y. Altmann *et al.*, Robust spectral unmixing of sparse multispectral Lidar waveforms using gamma Markov random fields. *IEEE Trans. Computat. Imaging* **3**, 658–670 (2017). doi: [10.1109/TCI.2017.2703144](https://doi.org/10.1109/TCI.2017.2703144)
79. P. Chhabra, A. Maccarone, A. McCarthy, G. S. Buller, A. Wallace, “Discriminating underwater Lidar target signatures using sparse multi-spectral depth codes” in *Sensor Signal Processing for Defence (SSPD)* (IEEE, 2016), pp. 1–5.
80. Y. Altmann, A. Maccarone, A. McCarthy, G. S. Buller, S. McLaughlin, “Joint spectral clustering and range estimation for 3D scene reconstruction using multispectral Lidar waveforms” in *26th European Signal Processing Conference (EUSIPCO 2016)* (2016), pp. 513–517.
81. Y. Altmann, A. Maccarone, A. McCarthy, G. Buller, S. McLaughlin, “Joint range estimation and spectral classification for 3D scene reconstruction using multispectral Lidar waveforms” in *2007 IEEE/SP 14th Workshop on Statistical Signal Processing* (IEEE, 2016), pp. 1–5.
82. Y. Altmann, A. Maccarone, A. McCarthy, S. McLaughlin, G. S. Buller, Spectral classification of sparse photon depth images. *Opt. Express* **26**, 5514–5530 (2018). doi: [10.1364/OE.26.005514](https://doi.org/10.1364/OE.26.005514); pmid: [29529755](https://pubmed.ncbi.nlm.nih.gov/29529755/)
83. R. Tobin *et al.*, Comparative study of sampling strategies for sparse photon multispectral lidar imaging: Towards mosaic filter arrays. *J. Opt.* **19**, 094006 (2017). doi: [10.1088/2040-8986/aa8237](https://doi.org/10.1088/2040-8986/aa8237)
84. Y. Altmann, R. Tobin, A. Maccarone, X. Ren, A. McCarthy, G. S. Buller, S. McLaughlin, Bayesian restoration of reflectivity and range profiles from subsampled single-photon multispectral Lidar data” in *EUSIPCO 2017* (IEEE, 2017), 1410–1414.
85. R. Takagi, R. Horisaki, J. Tanida, Object recognition through a multi-mode fiber. *Opt. Rev.* **24**, 117–120 (2017). doi: [10.1007/s10043-017-0303-5](https://doi.org/10.1007/s10043-017-0303-5)
86. A. Sinhai, J. Lee, S. Li, G. Barbastathis, Lensless computational imaging through deep learning. *Optica* **4**, 1117–1125 (2017). doi: [10.1364/OPTICA.4.001117](https://doi.org/10.1364/OPTICA.4.001117)
87. G. Satat, M. Tancik, O. Gupta, B. Heshmat, R. Raskar, Object classification through scattering media with deep learning on time resolved measurement. *Opt. Express* **25**, 17466–17479 (2017). doi: [10.1364/OE.25.017466](https://doi.org/10.1364/OE.25.017466); pmid: [28789238](https://pubmed.ncbi.nlm.nih.gov/28789238/)
88. P. Caramazza *et al.*, Neural network identification of people hidden from view with a single-pixel, single-photon detector. *arxiv:1709.07244* [cs.CV] (21 September 2017).
89. D. Shin, F. Xu, F. N. C. Wong, J. H. Shapiro, V. K. Goyal, Computational multi-depth single-photon imaging. *Opt. Express* **24**, 1873–1888 (2016). doi: [10.1364/OE.24.001873](https://doi.org/10.1364/OE.24.001873); pmid: [26906766](https://pubmed.ncbi.nlm.nih.gov/26906766/)
90. R. Tobin *et al.*, Long-range depth profiling of camouflaged targets using single-photon detection. *Opt. Eng.* **57**, 031303 (2017).
91. S. Medin, J. Murray-Bruce, V. K. Goyal, “Optimal stopping times for estimating Bernoulli parameters with applications to active imaging,” paper presented at *IEEE ICASSP*, Calgary, Alberta, Canada, 15 to 20 April 2018.

## ACKNOWLEDGMENTS

We thank the Royal Society for support and hosting the Theo Murphy Scientific meeting on “Light transport and imaging through complex media.” **Funding:** Y.A. acknowledges support from the UK Royal Academy of Engineering under the Research Fellowship Scheme (RF201617/16/31). S.McL. acknowledges financial support from the UK Engineering and Physical Sciences Research Council (grant EP/J015180/1). V.G. acknowledges support from the U.S. Defense Advanced Research Projects Agency (DARPA) InPho program through U.S. Army Research Office award W911NF-10-1-0404, the U.S. DARPA REVEAL program through contract HRO011-16-C-0030, and U.S. National Science Foundation through grants 1161413 and 1422034. A.H. acknowledges support from U.S. Army Research Office award W911NF-15-1-0479, U.S. Department of the Air Force grant FA8650-15-D-1845, and U.S. Department of Energy National Nuclear Security Administration grant DE-NA0002534. D.F. acknowledges financial support from the UK Engineering and Physical Sciences Research Council (grants EP/M006514/1 and EP/M01326X/1). **Competing interests:** None declared.

10.1126/science.aat2298

## RESEARCH ARTICLE SUMMARY

## WHEAT GENOME

# Shifting the limits in wheat research and breeding using a fully annotated reference genome

International Wheat Genome Sequencing Consortium (IWGSC)\*

**INTRODUCTION:** Wheat (*Triticum aestivum* L.) is the most widely cultivated crop on Earth, contributing about a fifth of the total calories consumed by humans. Consequently, wheat yields and production affect the global economy, and failed harvests can lead to social unrest. Breeders continuously strive to develop improved varieties by fine-tuning genetically complex yield and end-use quality parameters while maintaining stable yields and adapting the crop to regionally specific biotic and abiotic stresses.

**RATIONALE:** Breeding efforts are limited by insufficient knowledge and understanding of

wheat biology and the molecular basis of central agronomic traits. To meet the demands of human population growth, there is an urgent need for wheat research and breeding to accelerate genetic gain as well as to increase and protect wheat yield and quality traits. In other plant and animal species, access to a fully annotated and ordered genome sequence, including regulatory sequences and genome-diversity information, has promoted the development of systematic and more time-efficient approaches for the selection and understanding of important traits. Wheat has lagged behind, primarily owing to the challenges of assembling a genome that is more than five times as large

as the human genome, polyploid, and complex, containing more than 85% repetitive DNA. To provide a foundation for improvement through molecular breeding, in 2005, the International Wheat Genome Sequencing Consortium set out to deliver a high-quality annotated reference genome sequence of bread wheat.

**RESULTS:** An annotated reference sequence representing the hexaploid bread wheat genome in the form of 21 chromosome-like sequence assemblies has now been delivered, giving access to 107,891 high-confidence genes, including their genomic context of regulatory sequences. This assembly enabled the discovery of tissue- and developmental stage-related gene coexpression networks using a transcriptome atlas representing all stages of wheat development. The dynamics of change in complex gene

## ON OUR WEBSITE

Read the full article at <http://dx.doi.org/10.1126/science.aar7191>

families involved in environmental adaptation and end-use quality were revealed at subgenome resolution and contextualized to known agronomic single-gene or quantitative trait loci. As-

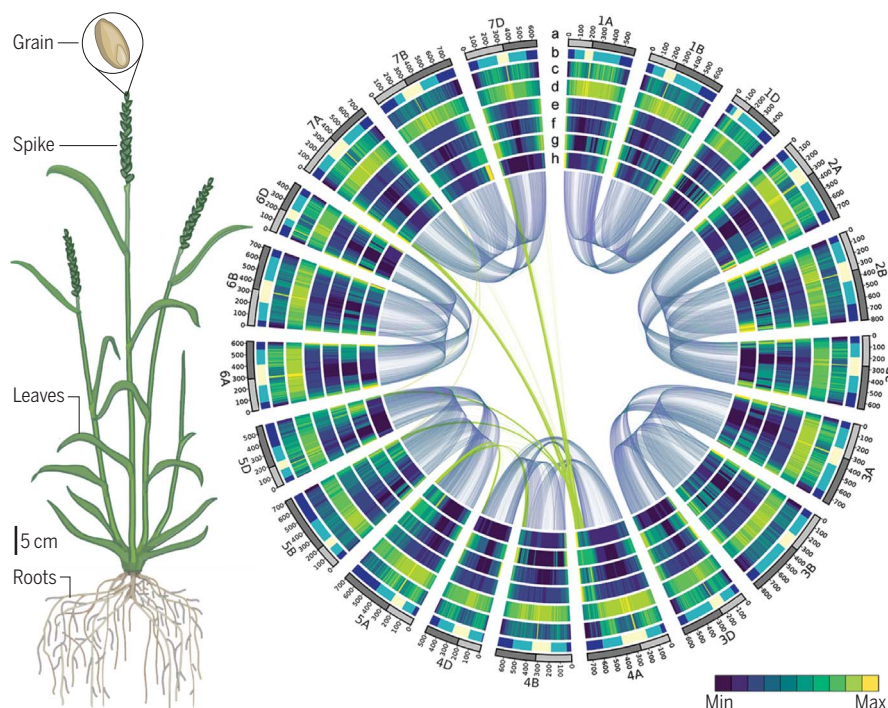
pects of the future value of the annotated assembly for molecular breeding and research were exemplarily illustrated by resolving the genetic basis of a quantitative trait locus conferring resistance to abiotic stress and insect damage as well as by serving as the basis for genome editing of the flowering-time trait.

**CONCLUSION:** This annotated reference sequence of wheat is a resource that can now drive disruptive innovation in wheat improvement, as this community resource establishes the foundation for accelerating wheat research and application through improved understanding of wheat biology and genomics-assisted breeding. Importantly, the bioinformatics capacity developed for model-organism genomes will facilitate a better understanding of the wheat genome as a result of the high-quality chromosome-based genome assembly. By necessity, breeders work with the genome at the whole chromosome level, as each new cross involves the modification of genome-wide gene networks that control the expression of complex traits such as yield. With the annotated and ordered reference genome sequence in place, researchers and breeders can now easily access sequence-level information to precisely define the necessary changes in the genomes for breeding programs. This will be realized through the implementation of new DNA marker platforms and targeted breeding technologies, including genome editing. ■

The list of author affiliations is available in the full article online.

\*Corresponding authors: Rudi Appels (rudi.appels@unimelb.edu.au); Kellye Eversole (eversole@eversoleassociates.com); Nils Stein (stein@ipk-gatersleben.de)

Cite this article as International Wheat Genome Sequencing Consortium, *Science* 361, eaar7191 (2018). DOI: 10.1126/science.aar7191



**Wheat genome deciphered, assembled, and ordered.** Seeds, or grains, are what counts with respect to wheat yields (left panel), but all parts of the plant contribute to crop performance. With complete access to the ordered sequence of all 21 wheat chromosomes, the context of regulatory sequences, and the interaction network of expressed genes—all shown here as a circular plot (right panel) with concentric tracks for diverse aspects of wheat genome composition—breeders and researchers now have the ability to rewrite the story of wheat crop improvement. Details on value ranges underlying the concentric heatmaps of the right panel are provided in the full article online.



## RESEARCH ARTICLE

## WHEAT GENOME

# Shifting the limits in wheat research and breeding using a fully annotated reference genome

International Wheat Genome Sequencing Consortium (IWGSC)\*

An annotated reference sequence representing the hexaploid bread wheat genome in 21 pseudomolecules has been analyzed to identify the distribution and genomic context of coding and noncoding elements across the A, B, and D subgenomes. With an estimated coverage of 94% of the genome and containing 107,891 high-confidence gene models, this assembly enabled the discovery of tissue- and developmental stage-related coexpression networks by providing a transcriptome atlas representing major stages of wheat development. Dynamics of complex gene families involved in environmental adaptation and end-use quality were revealed at subgenome resolution and contextualized to known agronomic single-gene or quantitative trait loci. This community resource establishes the foundation for accelerating wheat research and application through improved understanding of wheat biology and genomics-assisted breeding.

**W**heat (*Triticum aestivum* L.), the most widely cultivated crop on Earth, contributes about a fifth of the total calories consumed by humans and provides more protein than any other food source (1, 2). Breeders strive to develop improved varieties by fine-tuning genetically complex yield and end-use quality parameters while maintaining yield stability and regional adaptation to specific biotic and abiotic stresses (3). These efforts are limited, however, by insufficient knowledge and understanding of the molecular basis of key agronomic traits. To meet the demands of human population growth, there is an urgent need for wheat research and breeding to accelerate genetic gain while increasing wheat yield and protecting quality traits. In other plant and animal species, access to a fully annotated and ordered genome sequence, including regulatory sequences and genome-diversity information, has promoted the development of systematic and more time-efficient approaches for the selection and understanding of important traits (4). Wheat has lagged behind other species, primarily owing to the challenges of assembling a large (haploid genome, 1C = 16 Gb) (5), hexaploid, and complex genome that contains more than 85% repetitive DNA.

To provide a foundation for improvement through molecular breeding, the International Wheat Genome Sequencing Consortium (IWGSC) established a road map to deliver a high-quality reference genome sequence of the bread wheat cultivar Chinese Spring (CS). A chromosome survey sequence (CSS) intermediate product assigned 124,201 gene loci across the 21 chromosomes and revealed the evolutionary dynamics of

the wheat genome through gene loss, gain, and duplication (6). The lack of global sequence continuity and incomplete coverage (only 10 Gb were assembled), however, did not provide the wider regulatory genomic context of genes. Subsequent whole-genome assemblies improved contiguity (7–9) but lacked full annotation and did not resolve the intergenic space or present the genome in the correct physical order.

Here we report an ordered and annotated assembly (IWGSC RefSeq v1.0) of the 21 chromosomes of the allohexaploid wheat cultivar CS, an achievement that is built on a rich history of chromosome studies in wheat (10–12), which allowed the integration of genetic and genomic resources. The completeness and accuracy of IWGSC RefSeq v1.0 provide insights into global genome composition and enable the construction of complex gene coexpression networks to identify central regulators in critical pathways, such as flowering-time control. The ability to resolve the inherent complexity of gene families related to important agronomic traits demonstrates the impact of IWGSC RefSeq v1.0 on dissecting quantitative traits genetically and implementing modern breeding strategies for future wheat improvement.

## Chromosome-scale assembly of the wheat genome

Pseudomolecule sequences representing the 21 chromosomes of the bread wheat genome were assembled by integrating a draft de novo whole-genome assembly (WGA), built from Illumina short-read sequences using NRGene deNovoMagic2 (Fig. 1A, Table 1, and tables S1 and S2), with additional layers of genetic, physical, and sequence data (tables S3 to S8 and figs. S1 and S2). In the resulting 14.5-Gb genome assembly, contigs and

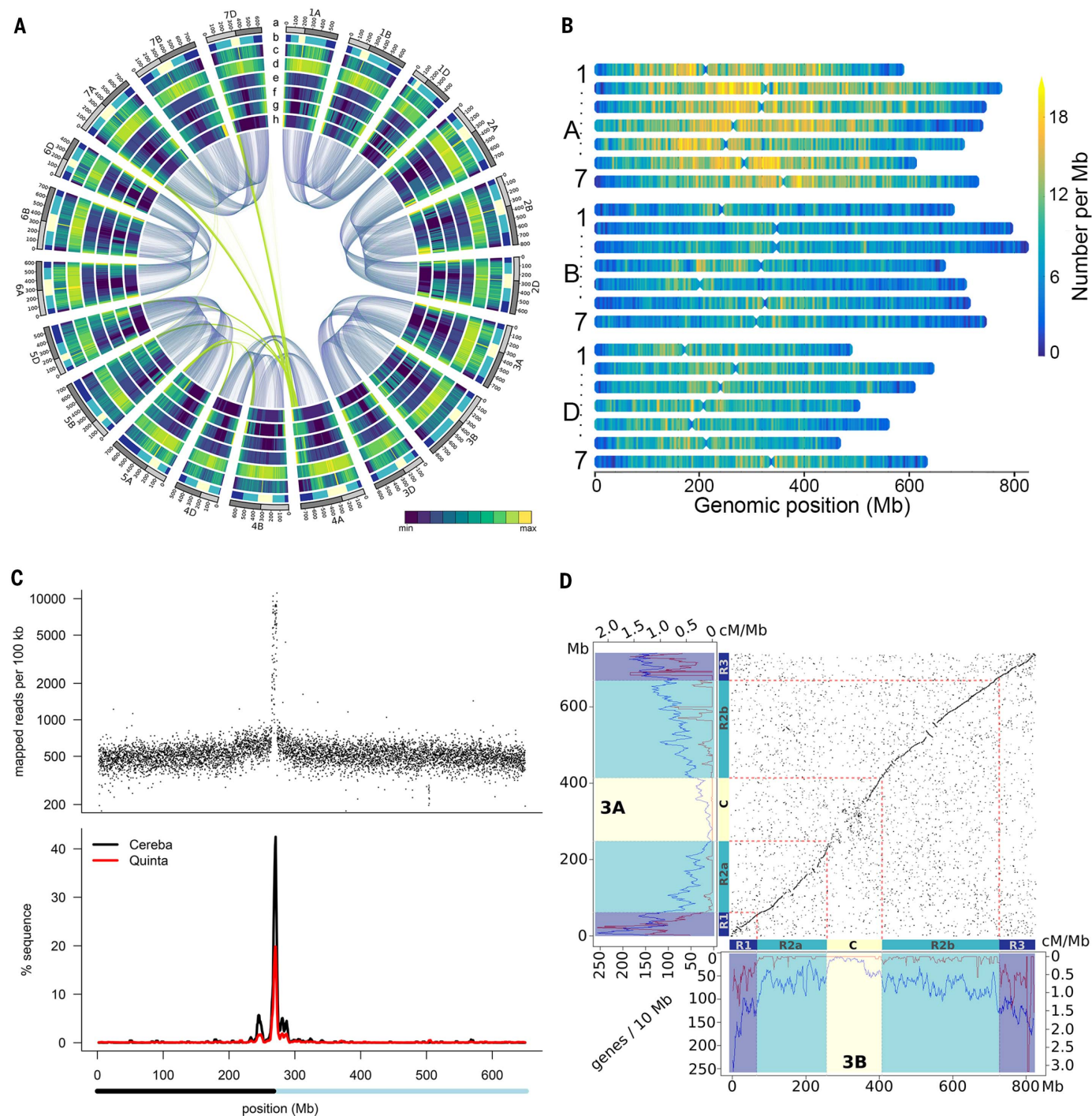
scaffolds with N50s of 52 kb and 7 Mb, respectively, were linked into superscaffolds (N50 = 22.8 Mb), with 97% (14.1 Gb) of the sequences assigned and ordered along the 21 chromosomes and almost all of the assigned sequence scaffolds oriented relative to each other (13.8 Gb, 98%). Unanchored scaffolds comprising 481 Mb (2.8% of the assembly length) formed the “unassigned chromosome” (ChrUn) bin. The quality and continuity of the IWGSC RefSeq v1.0 genome assembly were assessed through alignments with radiation hybrid maps for the A, B, and D subgenomes [average Spearman’s correlation coefficient ( $r$ ) of 0.98], the genetic positions of 7832 and 4745 genotyping-by-sequencing derived genetic markers in 88 double haploid and 993 recombinant inbred lines (Spearman’s  $r$  of 0.986 and 0.987, respectively), and 1.24 million pairs of neighbor insertion site-based polymorphism (ISBP) markers (13), of which 97% were collinear and mapped in a similar size range (difference of <2 kb) between the de novo WGA and the available bacterial artificial chromosome (BAC)-based sequence assemblies. Finally, IWGSC RefSeq v1.0 was assessed with independent data derived from coding and noncoding sequences, revealing that 99 and 98% of the previously known coding exons (6) and transposable element (TE)-derived (ISBP) markers (table S9), respectively, were present in the assembly. The approximate 1-Gb size difference between IWGSC RefSeq v1.0 and the new genome size estimates of 15.4 to 15.8 Gb (14) can be accounted for by collapsed or unassembled sequences of highly repeated clusters, such as ribosomal RNA coding regions and telomeric sequences.

A key feature distinguishing the IWGSC RefSeq v1.0 from previous draft wheat assemblies (6–9) is the long-range organization, with 90% of the genome represented in superscaffolds larger than 4.1 Mb and with each chromosome represented, on average, by only 76 superscaffolds (Table 1). The largest superscaffold spans 166 Mb, which is half the size of the rice (*Oryza sativa* L.) genome and is larger than the *Arabidopsis thaliana* L. genome (15, 16). Moreover, the 21 pseudomolecules position molecular markers for wheat research and breeding [504 single-stranded repeats (SSRs), 3025 diversity array technologies (DARs), 6689 expressed sequence tags (ESTs), 205,807 single-nucleotide polymorphisms (SNPs), and 4,512,979 ISBPs] (table S9), thus providing a direct link between the genome sequence and genetic loci and genes underlying traits of agronomic importance.

## The composition of the wheat genome

Analyses of the components of the genome sequence revealed the distribution of key elements and enabled detailed comparisons of the homeologous A, B, and D subgenomes. Accounting for 85% of the genome, with a relatively equal distribution across the three subgenomes (Table 2), 3,968,974 copies of TEs belonging to 505 families were annotated. Many (112,744) full-length long terminal repeat (LTR)-retrotransposons were identified that have been difficult to define from short-read sequence assemblies (fig. S3). Although the TE content has been extensively rearranged

\*All authors with their affiliations are listed at the end of this paper.



**Fig. 1. Structural, functional, and conserved synteny landscape of the 21 wheat chromosomes.** (A) Circular diagram showing genomic features of wheat. The tracks toward the center of the circle display (a) chromosome name and size (100-Mb tick size; light gray bar indicates the short arm and dark gray indicates the long arm of the chromosome); (b) dimension of chromosomal segments R1, R2a, C, R2b, and R3 [(18) and table S29]; (c) K-mer 20-frequencies distribution; (d) LTR-retrotransposons density; (e) pseudogenes density (0 to 130 genes per Mb); (f) density of HC gene models (0 to 32 genes per Mb); (g) density of recombination rate; and (h) SNP density. Connecting lines in the center of the diagram highlight homeologous relationships of chromosomes (blue lines) and translocated regions (green lines). (B) Distribution of Pfam domain

PF08284 “retroviral aspartyl protease” signatures across the different wheat chromosomes. (C) Positioning of the centromere in the 2D pseudomolecule. Top panel shows density of CENH3 ChIP-seq data along the wheat chromosome. Bottom panel shows distribution and proportion of the total pseudomolecule sequence composed of TEs of the Cereba and Quinta families. The bar below the bottom panel indicates pseudomolecule scaffolds assigned to the short (black) or long (blue) arm on the basis of CSS data [(6) mapping]. (D) Dot-plot visualization of collinearity between homeologous chromosomes 3A and 3B in relation to distribution of gene density and recombination frequency (left and bottom panel boxes: blue and purple lines, respectively). Chromosomal zones R1, R2a, C, R2b, and R3 are colored as in (A). cM, centimorgan.



through rounds of deletions and amplifications since the divergence of the A, B, and D subgenomes about 5 million years ago, the TE families that shaped the Triticeae genomes have been maintained in similar proportions: 76% of the 165 TE families present in a cumulative length greater than 1 Mb contributed similar proportions (less than a twofold difference between subgenomes), and only 11 families, accounting for 2% of total TEs, showed a higher than threefold difference between two subgenomes (17). TE abundance accounts, in part, for the size differences between subgenomes—for example, 64% of the 1.2-Gb size difference between the B and D subgenomes can be attributed to lower gypsy retrotransposon content. Low-copy DNA content (primarily unclassified sequences) also varied between subgenomes, accounting, for example, for 97 Mb of the 245-Mb size difference between A and B subgenomes (fig. S4). As reported (18), no evidence was found for a major burst of transposition after polyploidization. The independent evolution in the diploid lineages was reflected in differences in the specific composition of the A, B, and D subgenomes at the subfamily (variants) level, as evidenced by subgenome-specific over-

representation of individual transposon domain signatures (Fig. 1B). See (17) for a more detailed analysis of the TE content and its impact on the evolution of the wheat genome.

In addition to TEs, annotation of the intergenic space included noncoding RNAs. We identified eight new microRNA families (fig. S5 and table S10) and the entire complement of tRNAs (which showed an excess of lysine tRNAs, fig. S6). Around 8000 nuclear-inserted plastid DNA segments and 11,000 nuclear-inserted mitochondrial DNA segments representing, respectively, 5 and 17 Mb were also revealed by comparing the genome assembly with complete plastid and mitochondrial genomes assembled from the IWGSC RefSeq v1.0 raw read data (14).

Precise positions for the centromeres were defined by integrating Hi-C, CSS (6), and published chromatin immunoprecipitation sequencing (ChIP-seq) data for CENH3, a centromere-specific histone H3 variant (19). Clear ChIP-seq peaks were evident in all chromosomes and coincided with the centromere-specific repeat families (Fig. 1C, fig. S7, and table S11). CENH3 targets were also found in unassigned sequence scaffolds (ChrUn), indicating that centromeres of several

chromosomes are not yet completely resolved. On the basis of these data, a conservative estimate for the minimal average size of a wheat centromere is 4.9 Mb (6.7 Mb, if including ChrUn; table S11), compared with an average centromere size of ~1.8 Mb in maize (20, 21) and 0.4 to 0.8 Mb in rice (22).

Gene models were predicted with two independent pipelines previously utilized for wheat genome annotation and then consolidated to produce the RefSeq Annotation v1.0 (fig. S8). Subsequently, a set of manually curated gene models was integrated to build RefSeq Annotation v1.1 (fig. S9 and tables S12 to S17). In total, 107,891 high-confidence (HC) protein-coding loci were identified, with relatively equal distribution across the A, B, and D subgenomes (35,345, 35,643, and 34,212, respectively; Figs. 1D and 2A, fig. S10, and table S18). In addition, 161,537 other protein-coding loci were classified as low-confidence (LC) genes, representing partially supported gene models, gene fragments, and orphans (table S18). A predicted function was assigned to 82.1% (90,919) of HC genes in RefSeq Annotation v1.0 (tables S19 and S20), and evidence for transcription was found for 85% (94,114) of the HC genes versus 49% of the LC genes (23). Within the pseudogene category, 25,419 (8%) of 303,818 candidates matched LC gene models. The D subgenome contained significantly fewer pseudogenes than the A and B subgenomes (81,905 versus 99,754 and 109,097, respectively;  $\chi^2$  test,  $P < 2.2 \times 10^{-16}$ ) (tables S21 and S22 and fig. S10). In ChrUn, 2691 HC and 675 LC gene models were identified.

The quality of the RefSeq Annotation v1.1 gene set was benchmarked against BUSCO v3 (24), representing 1440 Embryophyta near-universal single-copy orthologs and published annotated wheat gene sets (Fig. 2B and fig. S11). Of the BUSCO v3 genes, 99% (1436) were represented in at least one complete copy in RefSeq Annotation v1.1 and 90% (1292) in three complete copies, an improvement over the 25% (353) and 70% (1014) of BUSCO v3 genes that were identified in the IWGSC (6) and TGACv1 (8) gene sets, respectively (Fig. 2B). Improved contiguity of sequences in the immediate vicinity of genes was also found: 61% of the HC and LC genes were flanked by at least 10 kb of sequence without ambiguous bases (Ns), in contrast to 37% and only 5% of the HC and LC genes in the TGACv1 and IWGSC CSS gene models, respectively (fig. S12).

To further characterize the gene space, a phylogenomic approach was applied to identify gene homeologs and paralogs between and within the wheat subgenomes and orthologs in other plant genomes (table S23 and figs. S13 to S15). Analysis of a subset of 181,036 genes [“filtered gene set,” (14) and Table 3] comprising 103,757 HC and 77,279 LC genes identified 39,238 homeologous groups—that is, clades of A, B, and D subgenome orthologs deduced from gene trees—containing a total of 113,653 genes (63% of the filtered set). Gene losses or retention and gene gains (gene duplications) were determined for all homeologous loci of IWGSC RefSeq v1.0 (Table 3), assuming the presence of a single gene copy at

Table 1. Assembly statistics of IWGSC RefSeq v1.0.

| Assembly characteristics                                   | Values                      |
|------------------------------------------------------------|-----------------------------|
| Assembly size                                              | 14.5 Gb                     |
| Number of scaffolds                                        | 138,665                     |
| Size of assembly in scaffolds $\geq$ 100 kb                | 14.2 Gb                     |
| Number of scaffolds $\geq$ 100 kb                          | 4,443                       |
| N50 contig length                                          | 51.8 kb                     |
| Contig L50 number                                          | 81,427                      |
| N90 contig length                                          | 11.7 kb                     |
| Contig L90 number                                          | 294,934                     |
| Largest contig                                             | 580.5 kb                    |
| Ns in contigs                                              | 0                           |
| N50 scaffold length                                        | 7.0 Mb                      |
| Scaffold L50 number                                        | 571                         |
| N90 scaffold length                                        | 1.2 Mb                      |
| Scaffold L90 number                                        | 2,390                       |
| Largest scaffold                                           | 45.8 Mb                     |
| Ns in scaffolds                                            | 261.9 Mb                    |
| Gaps filled with BAC sequences                             | 183 (1.7 Mb)                |
| Average size of inserted BAC sequence                      | 9.5 kb                      |
| N50 superscaffold length                                   | 22.8 Mb                     |
| Superscaffold L50 number                                   | 166                         |
| N90 superscaffold length                                   | 4.1 Mb                      |
| Superscaffold L90 number                                   | 718                         |
| Largest superscaffold                                      | 165.9 Mb                    |
| Sequence assigned to chromosomes                           | 14.1 Gb (96.8%)             |
| Sequence $\geq$ 100 kb assigned to chromosomes             | 14.1 Gb (99.1%)             |
| Number of superscaffolds on chromosomes                    | 1,601                       |
| Number of oriented superscaffolds                          | 1,243                       |
| Length of oriented sequence                                | 13.8 Gb (95%)               |
| Length of oriented sequence $\geq$ 100 kb                  | 13.8 Gb (97.3%)             |
| Smallest number of superscaffolds per subgenome chromosome | 35 (7A), 68 (2B), 36 (1D)   |
| Largest number of superscaffolds per subgenome chromosome  | 111 (4A), 176 (3B), 90 (3D) |
| Average number of superscaffolds per chromosome            | 76                          |

every homeologous locus (referred to as a “triad”). The percentage of genes in homeologous groups for all configurations (ratios) is highly similar, hence balanced, across the three subgenomes: 63% (A), 61% (B), and 66% (D). The slightly higher percentage of homeologs in the D subgenome, together with the lower number of pseudogenes (table S22), is consistent with its more recent hybridization with the AABB tetraploid genome progenitor. Although most of the genes are present in homeologous groups, only 18,595 (47%) of the groups contained triads with a single gene copy per subgenome (an A:B:D configuration of 1:1:1). Of the groups of homeologous genes, 5673 (15%) exhibited at least one subgenome inparalog, that is, a gene copy resulting from a tandem or a segmental trans duplication (1:1:N A:B:D configuration; N indicates a minimum of one additional paralog per respective subgenome). The three genomes exhibited similar levels of loss of individual homeologs, affecting 10.7% (0:1:1), 10.3% (1:0:1), and 9.5% (1:1:0) of the homeologous groups in the A, B, and D subgenomes, respectively (Table 3 and tables S24 and S25).

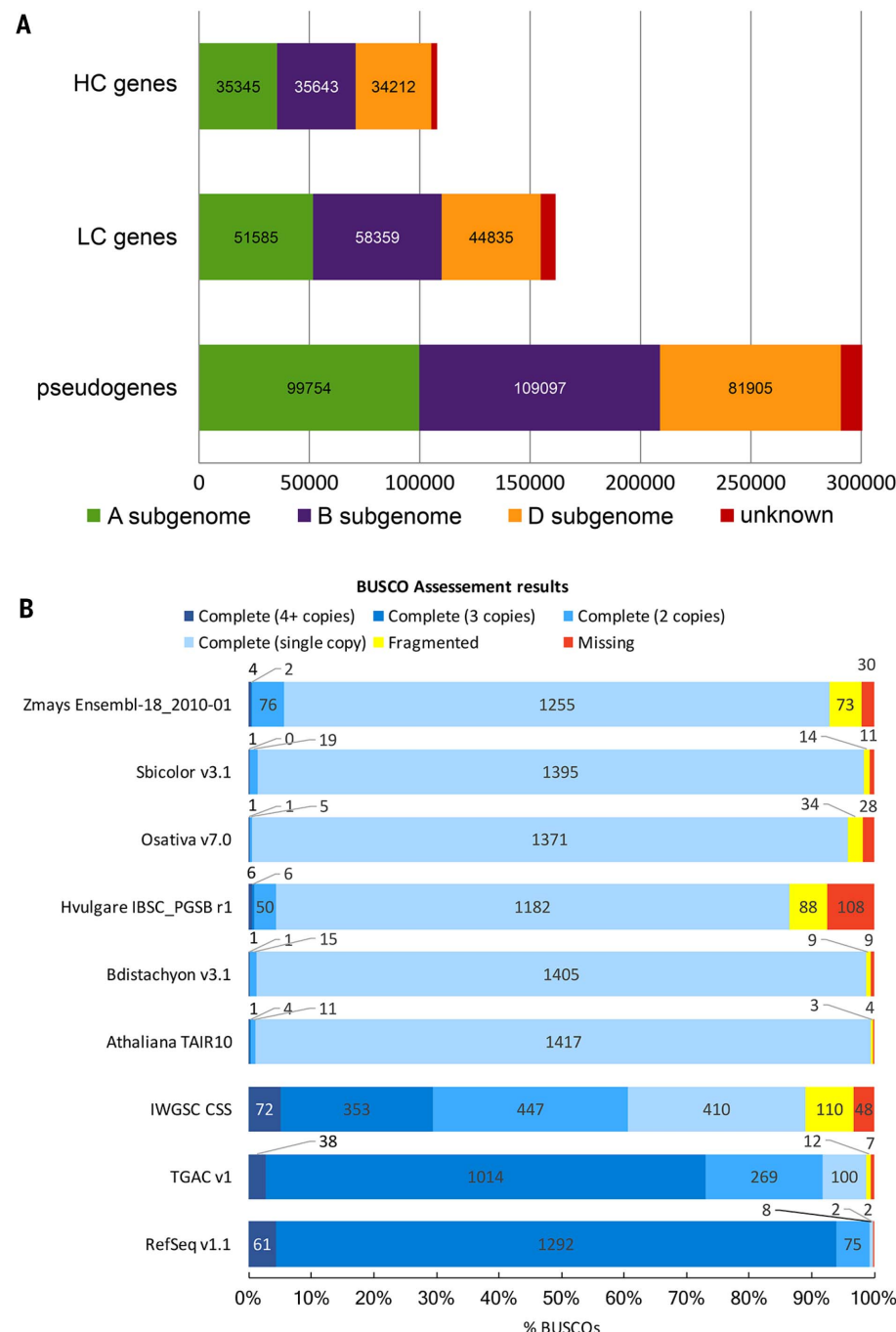
Of the 67,383 (37%) genes of the filtered set not present in homeologous groups, 31,140 genes also had no orthologs in species included in the comparisons outside of bread wheat and mainly comprised gene fragments, non-protein-coding loci with open reading frames, or other gene-calling artifacts. The remaining 36,243 genes had homologs outside of bread wheat and appeared to be subgenome specific (Table 3). Two of the genes in this category were *granule bound starch synthase (GBSS)* on chromosome 4A (1:0:0, a gene that is a key determinant of udon noodle quality) and *ZIP4* within the *pairing homeologous 1 (Ph1)* locus on chromosome 5B [0:1:0, a locus critical for the diploid meiotic behavior of the wheat homeologous chromosomes (25)]. The phylogenomic analysis indicated that the *GBSS* on 4A is a divergent translocated homeolog originally located on chromosome 7B (fig. S16), whereas *ZIP4* is a transduplication of a chromosome 3B locus (table S26). Both genes confer important properties on wheat and illustrate the diversity in origin and function of gene models that are not in a 1:1:1 configuration. No evidence was found for biased partitioning. Rather, our analyses support gradual gene loss and gene movement among the subgenomes that may have occurred in either the diploid progenitor species or the tetraploid ancestor or following the final hexaploidization event in modern bread wheat (Table 3 and figs. S24 and S25). Together with the equal contribution of the three homeologous genomes to the overall gene expression (23), this demonstrates the absence of subgenome dominance (26).

Of the bread wheat HC genes, 29,737 (27%) are present as tandem duplicates, which is up to 10% higher than that found for other monocotyledonous species (table S27). Tandemly repeated genes are most prevalent in the B subgenome (29%), contributing to its higher gene content and larger number of 1:N:1 homeologous groups (Table 3). The postulated hybrid origin of the D

subgenome, as a result of interspecific crossing with AABB tetraploid genome progenitors 1 to 2 million years after they diverged (27), is consistent with the synonymous substitution rates of homeologous gene pairs (fig. S17). Homeologous groups with gene duplicates in at least one subgenome (1:1:N, 1:N:1, or N:1:1) showed elevated evolutionary rates (for the subgenome

carrying the duplicate) as compared with strict 1:1:1 or 1:1 groups (figs. S18 to S22). Homeologs with recent duplicates also showed higher levels of expression divergence (fig. S23), consistent with gene and genome duplications acting as a driver of functional innovation (28, 29).

Analysis of synteny between the seven triplets of homeologous chromosomes showed high levels



**Fig. 2. Evaluation of automated gene annotation.** (A) Selected gene prediction statistics of IWGSC RefSeq Annotation v1.1, including number and subgenome distribution of HC and LC genes as well as pseudogenes. (B) BUSCO v3 gene model evaluation comparing IWGSC RefSeq Annotation v1.1 to earlier published bread wheat whole-genome annotations, as well as to annotations of related grass reference-genome sequences. BUSCO provides a measure for the recall of highly conserved gene models.



of conservation. There was no evidence that any major rearrangements occurred since the A, B, and D subgenomes diverged ~5 million years ago (Fig. 1D), although collinearity between homeologs was disturbed by inversions occurring, on average, every 74.8 Mb, involving blocks of 10 genes or more (mean gene number of 48.2 with a mean size of 10.5 Mb) (Fig. 1D and table S28). Macrosynteny was conserved across centromere (C) regions, but collinearity (microsynteny) broke down specifically in these recombination-free, gene-poor regions for all seven sets of homeologous chromosomes (Fig. 1D, figs. S24 to S26, and table S29). Of the 113,653 homeologous genes, 80% (90,232) were found organized in macrosynteny, that is, still present at their ancestral position (table S24). At the microsynteny scale, 72% (82,308) of the homeologs were organized in collinear blocks, that is, intervals with a highly conserved gene order (Fig. 1D). A higher proportion of syntenic genes was found in the interstitial regions [short arm, R2a (18), 46% and long arm, R2b (18), 61%] than in the distal telomeric [short arm, R1 (18), 39% and long arm, R3 (18), 51%] and centromere regions [C (18), 29%], and the interstitial compartments harbored larger syntenic blocks (figs. S27 and S28). The higher proportions of duplicated genes in distal-terminal regions (34 and 27% versus 13 to 15% in the other

regions; fig. S29) exerted a strong influence on the decay of syntenic block size and contributed to the higher sequence variability in these regions. Overall, distal chromosomal regions are the preferential targets of meiotic recombination and the fastest evolving compartments. As such, they represent the genomic environment for creating sequence, hence allelic, diversity, providing the basis for adaptability to changing environments.

Atlas of transcription reveals trait-associated gene co-regulation networks

The gene annotation, coupled with identification of homeologs and paralogs in IWGSC RefSeq v1.0, provides a resource to study gene expression in genome-wide and subgenome contexts. A total of 850 RNA-seq samples derived from 32 tissues at different growth stages and/or challenged by different stress treatments were mapped to RefSeq Annotation v1.0 (Fig. 3A, database S1, and tables S30 to S32). Expression was observed for 94,114 (84.9%) HC genes (fig. S30) and for 77,920 (49.1%) LC genes, the latter showing lower expression breadth and level [median six tissues; average 2.9 transcripts per million (tpm)] than the HC genes (median 20 tissues, average 8.2 tpm) (fig. S31). This correlated with the higher average methylation status of LC genes (figs. S32 and S33). A principal component analysis identified tissue

(Fig. 3B), rather than growth stage or stress (fig. S34), as the main factor driving differential expression between samples, consistent with studies in other organisms (30–33). Of the total number of genes, 31.0% are expressed in more than 90% of tissues (average 16.9 tpm, ≥30 tissues), and 21.5% are expressed in 10% or fewer tissues (average 0.22 tpm, ≤3 tissues; fig. S31).

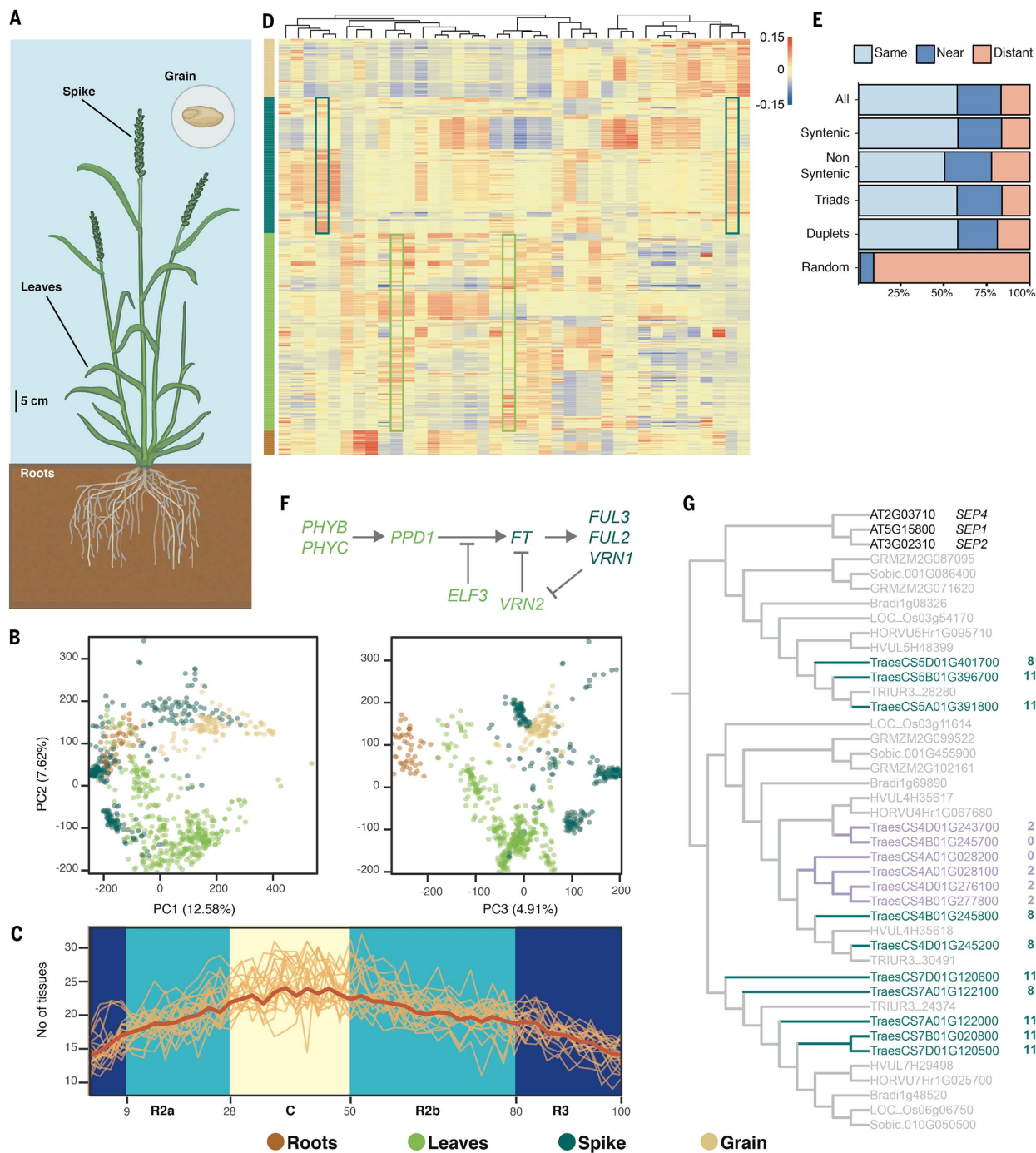
Of the HC genes, 8231 showed tissue-exclusive expression (fig. S35). About half of these were associated with reproductive tissues (microspores, anther, and stigma or ovary), consistent with observations in rice (34). The tissue-exclusive genes were enriched for response to extracellular stimuli and reproductive processes (database S2). By contrast, 23,146 HC genes expressed across all 32 tissues were enriched for biological processes associated with housekeeping functions such as protein translation and protein metabolic processes. Tissue-specific genes were shorter [1147 ± 8 base pairs (bp)], had fewer exons (2.76 ± 0.3), and were expressed at lower levels (3.4 ± 0.1 tpm) compared with ubiquitous genes (1429 ± 7 bp, 7.87 ± 0.4 exons, and 17.9 ± 0.4 tpm) (fig. S35).

Genes located in distal regions R1 and R3 (fig. S25 and table S29) showed lower expression breadth than those in the proximal regions (15.7 and 20.7 tissues, respectively) (Fig. 3C and fig. S36). This correlated with enrichment of Gene Ontology (GO) slim terms such as “cell cycle,” “translation,” and “photosynthesis” for genes in the proximal regions, whereas genes enriched for “response to stress” and “external stimuli” were found in the highly recombinant distal R1 and R3 regions (database S3, fig. S36, and table S33). The expression breadth pattern was also correlated with the distribution of the repressive H3K27me3 (trimethylated histone H3 lysine 27) (Pearson  $r = -0.76$ ,  $P < 2.2 \times 10^{-16}$ ) and with the active H3K36me3 and H3K9ac (acetylated H3K9) (Pearson  $r = 0.9$  and  $0.83$ , respectively;  $P < 2.2 \times 10^{-16}$ ) histone marks (fig. S37).

Global patterns of coexpression (35) were determined with a weighted gene coexpression network analysis (WGCNA) on 94,114 expressed HC genes. Of these genes, 58% (54,401) could be assigned to 38 modules (Fig. 3D and database S4), and, consistent with the principal component analysis, tissues were the major driver of module identity (Fig. 3D and figs. S38 to S40). The analysis focused initially on the 9009 triads (syntenic and nonsyntenic) with a 1:1:1 A:B:D relationship and for which all homeologs were assigned to a module. Of the triads, 16.4% had at least one homeolog in a divergent module, with the B homeolog most likely to be divergent (37.4% B-divergent versus 31.7% A-divergent and 30.9% D-divergent triads,  $\chi^2$  test  $P = 0.007$ ). However, the expression profiles of most (83.6%) of the triads were relatively consistent with all homeologs in the same (57.6%) or a closely related (26.0%) module. The proportion of homeologs found within the same module was higher than expected, pointing to a highly conserved expression pattern of homeologs across the 850 RNA-seq samples (Fig. 3E and table S34). Triads with at least one gene in a nonsyntenic position had a higher amount of

**Table 2. Relative proportions of the major elements of the wheat genome.** Proportions of TEs are given as the percentage of sequences assigned to each superfamily relative to genome size. Abbreviations in parentheses under the headings “Class 1” and “Class 2” indicate transposon types.

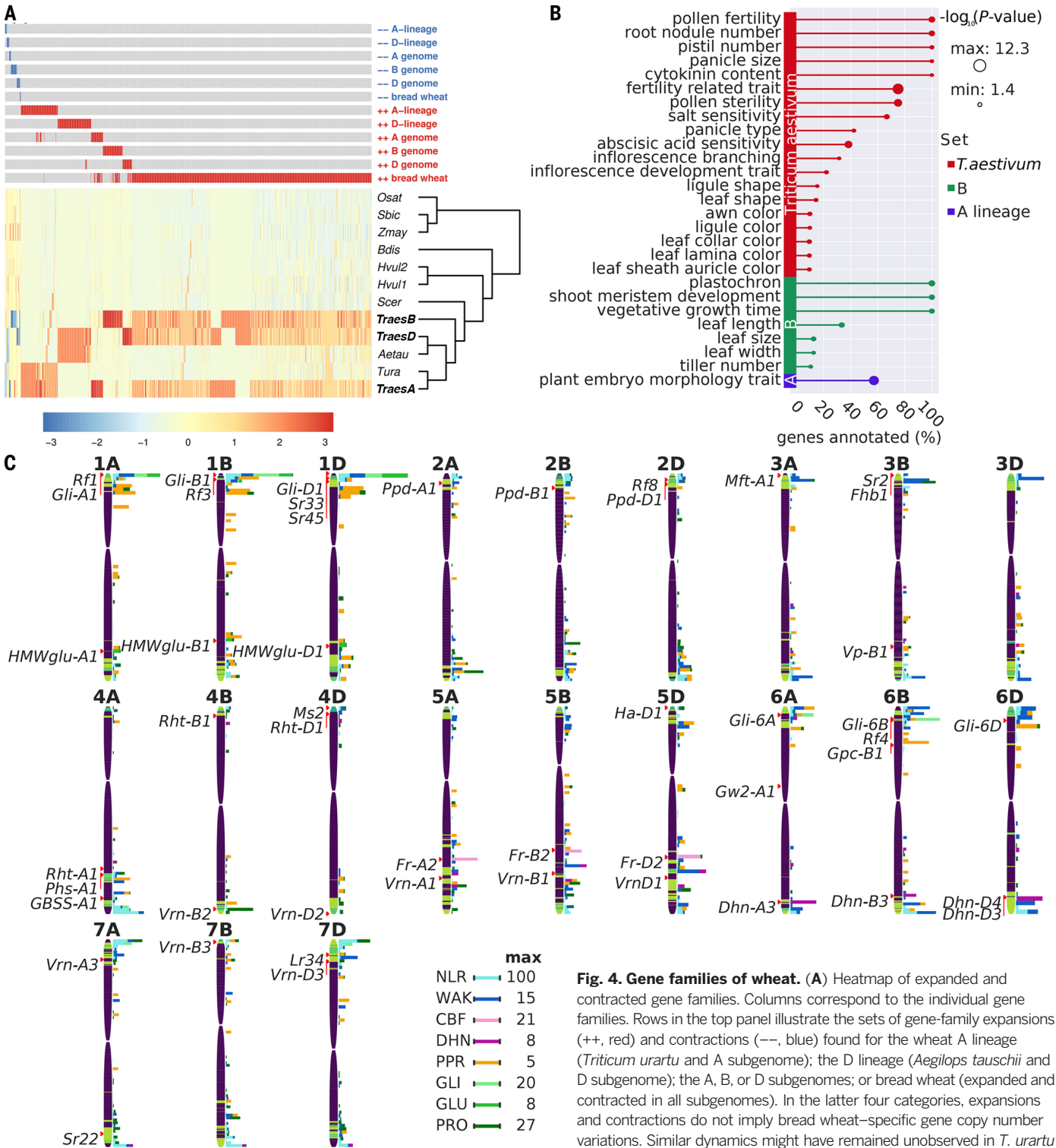
| Major elements                                  | Wheat subgenome |        |        |        |
|-------------------------------------------------|-----------------|--------|--------|--------|
|                                                 | AA              | BB     | DD     | Total  |
| Assembled sequence assigned to chromosomes (Gb) | 4.935           | 5.180  | 3.951  | 14.066 |
| Size of TE-related sequences (Gb)               | 4.240           | 4.388  | 3.285  | 11.913 |
| TEs (%)                                         | 85.9            | 84.7   | 83.1   | 84.7   |
| Class 1                                         |                 |        |        |        |
| LTR-retrotransposons                            |                 |        |        |        |
| Gypsy (RLG)                                     | 50.8            | 46.8   | 41.4   | 46.7   |
| Copia (RLC)                                     | 17.4            | 16.2   | 16.3   | 16.7   |
| Unclassified LTR-retrotransposons (RLX)         | 2.6             | 3.5    | 3.7    | 3.2    |
| Non-LTR-retrotransposons                        |                 |        |        |        |
| Long interspersed nuclear elements (RIX)        | 0.81            | 0.96   | 0.93   | 0.90   |
| Short interspersed nuclear elements (SIX)       | 0.01            | 0.01   | 0.01   | 0.01   |
| Class 2                                         |                 |        |        |        |
| DNA transposons                                 |                 |        |        |        |
| CACTA (DTC)                                     | 12.8            | 15.5   | 19.0   | 15.5   |
| Mutator (DTM)                                   | 0.30            | 0.38   | 0.48   | 0.38   |
| Unclassified with terminal inverted repeats     | 0.21            | 0.20   | 0.22   | 0.21   |
| Harbinger (DTH)                                 | 0.15            | 0.16   | 0.18   | 0.16   |
| Mariner (DTT)                                   | 0.14            | 0.16   | 0.17   | 0.16   |
| Unclassified class 2                            | 0.05            | 0.08   | 0.05   | 0.06   |
| hAT (DTA)                                       | 0.01            | 0.01   | 0.01   | 0.01   |
| Helitrons (DHH)                                 | 0.0046          | 0.0044 | 0.0036 | 0.0042 |
| Unclassified repeats                            | 0.55            | 0.85   | 0.63   | 0.68   |
| Coding DNA                                      | 0.89            | 0.89   | 1.11   | 0.95   |
| Unannotated DNA                                 | 13.2            | 14.4   | 15.7   | 14.4   |
| (Pre)-microRNAs                                 | 0.039           | 0.057  | 0.046  | 0.047  |
| tRNAs                                           | 0.0056          | 0.0050 | 0.0068 | 0.0057 |



**Fig. 3. Wheat atlas of transcription.** (A) Schematic illustration of a mature wheat plant and high-level tissue definitions for “roots,” “leaves,” “spike,” and “grain” used in the further analysis. (B) Principal component (PC) analysis plots for similarity of overall transcription, with samples colored according to their high-level tissue of origin [as introduced in (A)]. The color key for tissue is shown at the bottom of the figure under (C). (C) Chromosomal distribution of the average expression breadth [number of tissues in which genes are expressed (total number of tissues,  $n = 32$ )]. The average (dark orange line) is calculated on the basis of a scaled position of each gene within the corresponding genomic compartment (blue, aqua, and light yellow background) across the 21 chromosomes (orange lines). (D) Heatmap illustrating the expression of a representative gene (eigengene) for the 38 coexpression modules defined by WGCNA. Modules are represented as columns, with the dendrogram illustrating eigengene relatedness. Each row

represents one sample. Colored bars to the left indicate the high-level tissue of origin; the color key is shown at the bottom of the figure under (C). DESeq2-normalized expression levels are shown. Modules 1 and 5 (light green boxes) were most correlated with high-level leaf tissue, whereas modules 8 and 11 (dark green boxes) were most correlated with spike. (E) Bar plot of module assignment (same, near, or distant) of homeologous triads and duplets in the WGCNA network. (F) Simplified flowering pathway in polyploid wheat. Genes are colored according to their assignment to leaf (light green)– or spike (dark green)–correlated modules. (G) Excerpt from phylogenetic tree for MADS transcription factors, including known *Arabidopsis* flowering regulators *SEP1*, *SEP2*, and *SEP4* (black) (for the full phylogenetic tree, see fig. S38). Green branches represent wheat orthologs of modules 8 and 11, whereas purple branches are wheat orthologs assigned to other modules (0 and 2). Gray branches indicate non-wheat genes.





**Fig. 4. Gene families of wheat.** (A) Heatmap of expanded and contracted gene families. Columns correspond to the individual gene families. Rows in the top panel illustrate the sets of gene-family expansions (++, red) and contractions (—, blue) found for the wheat A lineage (*Triticum urartu* and A subgenome); the D lineage (*Aegilops tauschii* and D subgenome); the A, B, or D subgenomes; or bread wheat (expanded and contracted in all subgenomes). In the latter four categories, expansions and contractions do not imply bread wheat-specific gene copy number variations. Similar dynamics might have remained unobserved in *T. urartu* or *A. tauschii* owing to the inherent limitations of the used draft genome

assemblies (53, 54). Rows in the bottom panel heatmap (color scheme on z-score scale) indicate the fold expansion and contraction of gene families for the taxa and species included in the analysis [*Oryza sativa* (Osat), *Sorghum bicolor* (Sbic), *Zea mays* (Zmay), *Brachypodium distachyon* (Bdis), *Hordeum vulgare* (Hvul1/2), *Secale cereale* (Scer), *A. tauschii* (Aetau), *T. urartu* (Tura), and wheat A (TraesA), B (TraesB), and D (TraesD) subgenomes]. (B) All enriched TO terms for the gene families depicted in (A). Overrepresented TO terms were found for expanded families in bread wheat (all subgenomes, red), the B subgenome (green), and the A lineage (*T. urartu* and A subgenome, blue) only, respectively. The x axis represents the percentage of genes annotated with the respective TO term that were contained in the gene set in question. The size of the bubbles corresponds to the  $P$  ( $-\log_{10}$ ) significance of expansion. (C) Genomic distribution of gene families associated with adaptation to biotic (light and dark blue) or abiotic stress (light and dark pink), RNA metabolism in organelles and male fertility (orange), or end-use quality (light, medium, and dark green). Known positions of agronomically important genes and loci are indicated by red arrows and arrowheads to the left of the chromosome bars. Recombination rates are displayed as heatmaps in the chromosome bars [7.2 cM/Mb (light green) to 0 cM/Mb (black)].

divergent expression patterns compared to syntenic triads (21.2 versus 16.2%,  $\chi^2$  test  $P < 0.001$ ) and fewer such triads shared all homeologs in the same module (48.7%) compared to syntenic triads (58.0%, chi-square test  $P = 0.009$ ). Similar patterns were observed in the 1933 duplets that have a 1:1 relationship between only two homeologs (table S34). These results are consistent with syntenic homeologs showing similar expression patterns, whereas more dramatic changes in chromosome context associate with divergent expression and possible sub- or neofunctionalization. These trends were also found across diverse tissue-specific networks (23).

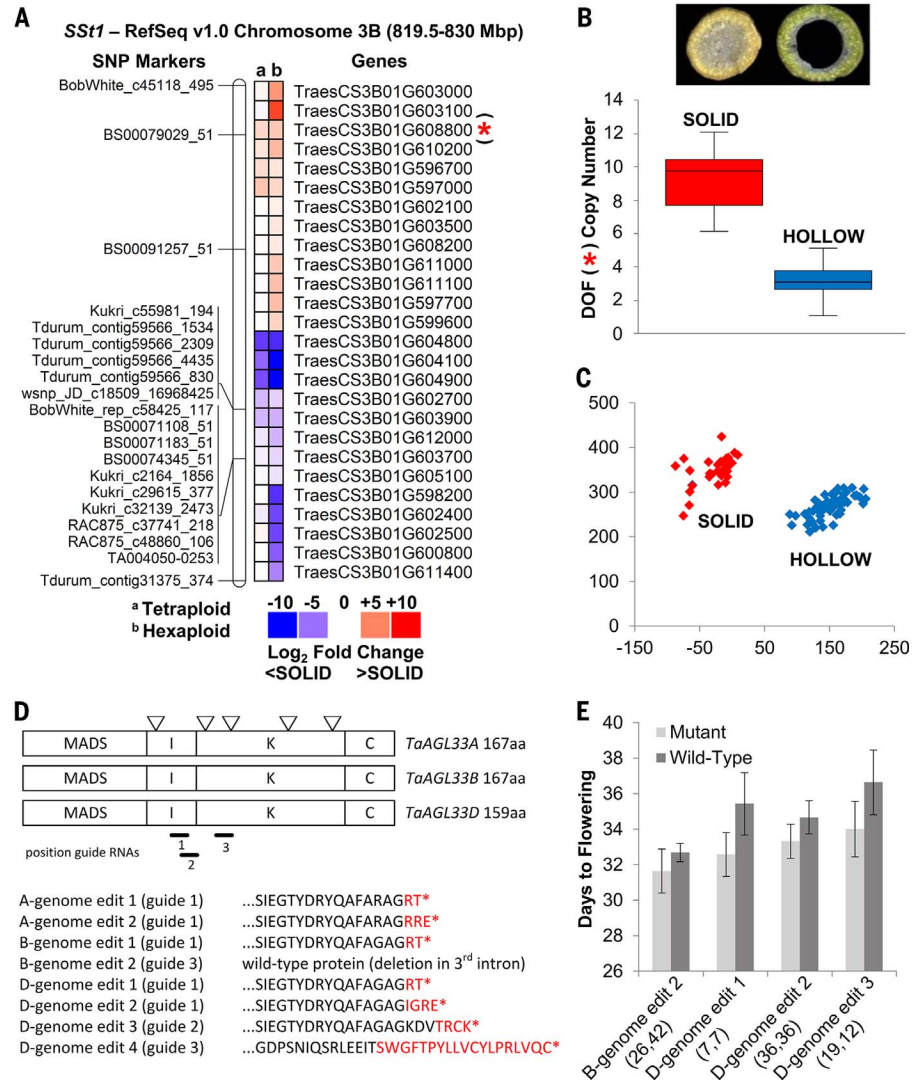
To explore the potential of the WGCNA network for identifying previously uncharacterized pathways in wheat, a search was undertaken for modules containing known regulators of wheat flowering time [e.g., *PPD1* (36) and *FT* (37); Fig. 3F]. Genes belonging to this pathway were grouped into specific modules. The upstream genes (*PHYB*, *PHYC*, *PPD1*, *ELF3*, and *VRN2*) were present mainly in modules 1 and 5 and were most highly correlated with expression in leaf and shoot tissues (0.68 and 0.67, respectively; adjusted  $P < 1 \times 10^{-108}$ ). By contrast, the integrating gene *FT* and downstream genes *VRN1*, *FUL2*, and *FUL3* were found in modules 8 and 11, most highly correlated with expression in spikes (0.69 and 0.65, respectively; adjusted  $P < 1 \times 10^{-101}$ ; table S35). The MADS\_II transcription factor family that is generally associated with the above pathways was examined more closely, with a focus on the gene tree OG0000041, which contains 54 of the 118 MADS\_II genes in wheat. Twenty-four MADS\_II genes from modules 8 and 11 were identified within this gene tree, clustering into two main clades along with *Arabidopsis* and rice orthologs associated with floral patterning (fig. S41 and database S5). Within these clades, other MADS\_II genes were found that were not in modules 8 or 11 (Fig. 3G), indicating a different pattern of coexpression. None of the 24 MADS\_II genes had a simple 1:1 ortholog in *Arabidopsis*, suggesting that some wheat orthologs function in flowering (those within modules 8 and 11), whereas others could have developed different functions, despite being phylogenetically closely related. Thus, these data provide a framework to identify and prioritize the most likely functional orthologs of known model system genes within polyploid wheat, to characterize them functionally (38), and to dissect genetic factors controlling important agronomic traits (39, 40). A more detailed analysis of tissue-specific and stress-related networks (23) provides a framework for defining quantitative variation and interactions between homeologs for many agronomic traits (41).

Gene-family expansion and contraction with relevance to wheat traits

Gene duplication and gene-family expansion are important mechanisms of evolution and environmental adaptation, as well as major contributors to phenotypic diversity (42, 43). In a phylogenomic comparative analysis, wheat gene-family

size and wheat-specific gene-family expansion and contraction were benchmarked against nine other grass genomes, including five closely related diploid Triticeae species (table S23 and figs. S13

to S15 and S42). A total of 30,597 gene families (groups of orthologous genes traced to a last common ancestor in the evolutionary hierarchy of the compared taxa) were defined, with 26,080



**Fig. 5. IWGSC RefSeq v1.0-guided dissection of *Sst1* and *TaAGL33*.** (A) The Lillian-Vesper population genetic map was anchored to IWGSC RefSeq v1.0 (left), and differentially expressed genes were identified between solid- and hollow-stemmed lines of hexaploid (bread) and tetraploid (durum) wheat (right). (B) Cross-sectioned stems of Lillian (solid) and Vesper (hollow) are shown as a phenotypic reference (top). Increased copy number of *TraesCS3B01G608800* [annotated as a DOF (DNA-binding one-zinc finger) transcription factor] is associated with stem phenotypic variation (bottom). (C) A high-throughput SNP marker tightly linked to *TraesCS3B01G608800* reliably discriminates solid- from hollow-stemmed wheat lines. Relative intensity of the fluorophores (FAM and HEX) used in KASPar analysis are shown. Vertical axis shows FAM signal; horizontal axis shows HEX signal. (D) Schematic of the three *TaAGL33* proteins, showing the typical MADS, I, K, and C domains. Triangles indicate the position of the five introns that occur in all three homeologs. Bars indicate the position of single-guide RNAs designed for exons 2 and 3. Three T-DNA vectors—each containing the *bar* selectable marker gene, CRISPR nuclease, and one of three single-guide RNA sequences—were used for *Agrobacterium*-mediated wheat transformation, essentially as described earlier (55). Transgenic plants were obtained with edits at the targeted positions in all *TaAGL33* homeologs. The putatively resulting protein sequence is displayed starting close to the edits, with wild-type amino acids (aa) in black font and amino acids resulting from the induced frame shifts in red font. \* indicates premature termination codons. (E) Mean days to flowering (after 8 weeks of vernalization) for progeny of four homozygous edited plants (light gray bars) and the respective homozygous wild-type segregants (dark gray bars). Numbers in parentheses refer to the number of edited and wild-type plants examined, respectively. Error bars display SEM. Growth conditions were as described in (50).



families containing gene members from at least one of the three wheat subgenomes (tables S36 to S39). Among the 8592 expanded wheat gene families (33% of all families), 6216 were expanded in all three A, B, and D subgenomes (24%; either shared with the wild ancestor or specific to bread wheat, Fig. 4A). Another 1109 were expanded in only one of the wheat subgenomes, and 2102 gene families were expanded in either the A or the D genome lineages (Fig. 4A, fig. S43, and table S36). Overall, only 78 gene families were contracted in wheat. The number of gene families that are only expanded in wheat may be overestimated owing to limited completeness of the draft progenitor wheat genome assemblies used in this study (14) (table S39). Gene Ontology (GO; ontology of biomedical terms for the areas “cellular component,” “biological process,” and “molecular function”), Plant Ontology (PO; ontology terms describing anatomical structures and growth and developmental stages across Viridiplantae), and Plant Trait Ontology [TO; ontology of controlled vocabulary to describe phenotypic traits and quantitative trait loci (QTLs) that were physically mapped to a gene in flowering plant species] analyses identified 1169 distinct GO, PO, and TO terms (15% of all assigned terms) enriched in genes belonging to expanded wheat gene families (Fig. 4B and figs. S44 and S45). “A-subgenome” or “A-lineage” expanded gene families showed a bias for terms associated with seed formation [overrepresentation of the TO term “plant embryo morphology” (TO:0000064) and several seed, endosperm, and

embryo-developmental GO terms] (fig. S46). Similarly, “B-subgenome” expanded gene families were enriched for TO terms related to plant vegetative growth and development (database S6 and fig. S47). Gene families that were expanded in all wheat subgenomes were enriched for 14 TO terms associated with yield-affecting morphological traits and five terms associated with fertility and abiotic-stress tolerance (Fig. 4B), which was also mirrored by enrichment for GO and PO terms associated with adaptation to abiotic stress (“salt stress” and “cold stress”) and grain yield and quality (“seed maturation,” “dormancy,” and “germination”). The relationship between the patterns of enriched TO, PO, and GO terms for expanded wheat gene families and key characteristics of wheat performance (figs. S45 to S51) provides a resource (database S6) to explore future QTL mapping and candidate gene identification for breeding.

Many gene families with high relevance to wheat breeding and improvement were among the expanded group, and their genomic distribution was analyzed in greater detail (Fig. 4C and figs. S52 to S54). Disease resistance-related NLR (nucleotide-binding site leucine-rich repeat)-like loci and WAK (wall-associated receptor)-like genes were clustered in high numbers at the distal (R1 and R3) regions of all chromosome arms, with NLRs often co-localizing with known disease resistance loci (Fig. 4C). The restorer-of-fertility-like (RFL) subclade of P class pentatricopeptide repeat (PPR) proteins, potentially of interest for hybrid wheat production, com-

prised 207 genes, nearly threefold more per haploid subgenome than have been identified in any other plant genome analyzed to date (44, 45). They localized mainly as clusters of genes in regions on the group 1, 2, and 6 chromosomes, which carry fertility-restoration QTLs in wheat (Fig. 4C and fig. S54). Within the dehydrin gene family, implicated with drought tolerance in plants, 25 genes that formed well-defined clusters on chromosomes 6A, 6B, and 6D (figs. S53 and S55) showed early increased expression under severe drought stress. As the structural variation in the *CBF* genes of wheat is known to be associated with winter survival (46), the array of *CBF* paralogs at the *Fr-2* locus (fig. S56) revealed by IWGSC RefSeq v1.0 provides a basis for targeted allele mining for previously uncharacterized *CBF* haplotypes from highly frost-tolerant wheat genetic resources. Lastly, high levels of expansion and variation in members of grain prolamin gene families [fig. S52 and (47)] either related to the response to heat stress or whose protein epitopes are associated with levels of celiac disease and food allergies (47) provide candidates for future selection in breeding programs. From these few examples, it is evident that flexibility in gene copy numbers within the wheat genome has contributed to the adaptability of wheat to produce high-quality grain in diverse climates and environments (48). Knowledge of the complex picture of the genome-wide distribution of gene families (Fig. 4C), which needs to be considered for selection in breeding programs in the context of

**Table 3. Groups of homeologous genes in wheat.** Homeologous genes are “subgenome orthologs” and were inferred by species tree reconciliation in the respective gene family. Numbers include both HC and LC genes filtered for TEs (filtered gene set). Conserved subgenome-specific (orphan) genes are found only in one subgenome but have homologs in other plant genomes used in this study. This includes orphan outparalogs resulting from ancestral duplication events and conserved only in one of the subgenomes. Nonconserved orphans

are either singletons or duplicated in the respective subgenome, but neither have obvious homologs in the other subgenomes or the other plant genomes studied. Microsynteny is defined as the conservation and collinearity of local gene ordering between orthologous chromosomal regions. Macrosynteny is defined as the conservation of chromosomal location and identity of genetic markers like homeologs but may include the occurrence of local inversions, insertions, or deletions. Additional data are presented in table S24.

| Homeologous group (A:B:D)                 | Number in wheat genome | Composition of groups (%) | Number of genes in A | Number of genes in B | Number of genes in D | Total number of genes |
|-------------------------------------------|------------------------|---------------------------|----------------------|----------------------|----------------------|-----------------------|
| 1:1:1                                     | 21,603                 | 55.1                      | 21,603               | 21,603               | 21,603               | 64,809                |
| 1:1:N                                     | 644                    | 1.6                       | 644                  | 644                  | 1,482                | 2,770                 |
| 1:N:1                                     | 998                    | 2.5                       | 998                  | 2,396                | 998                  | 4,392                 |
| N:1:1                                     | 761                    | 1.9                       | 1,752                | 761                  | 761                  | 3,274                 |
| 1:1:0                                     | 3,708                  | 9.5                       | 3,708                | 3,708                | 0                    | 7,416                 |
| 1:0:1                                     | 4,057                  | 10.3                      | 4,057                | 0                    | 4,057                | 8,114                 |
| 0:1:1                                     | 4,197                  | 10.7                      | 0                    | 4,197                | 4,197                | 8,394                 |
| Other ratios                              | 3,270                  | 8.3                       | 4,999                | 5,371                | 4,114                | 14,484                |
| 1:1:1 in microsynteny                     | 18,595                 | 47.4                      | 18,595               | 18,595               | 18,595               | 55,785                |
| Total in microsynteny                     | 30,339                 | 77.3                      | 27,240               | 27,063               | 28,005               | 82,308                |
| 1:1:1 in macrosynteny                     | 19,701                 | 50.2                      | 19,701               | 19,701               | 19,701               | 59,103                |
| Total in macrosynteny                     | 32,591                 | 83.1                      | 29,064               | 30,615               | 30,553               | 90,232                |
| <b>Total in homeologous groups</b>        | <b>39,238</b>          | <b>100.0</b>              | <b>37,761</b>        | <b>38,680</b>        | <b>37,212</b>        | <b>113,653</b>        |
| Conserved subgenome orphans               |                        |                           | 12,412               | 12,987               | 10,844               | 36,243                |
| Nonconserved subgenome singletons         |                        |                           | 10,084               | 12,185               | 8,679                | 30,948                |
| Nonconserved subgenome duplicated orphans |                        |                           | 71                   | 83                   | 38                   | 192                   |
| <b>Total (filtered)</b>                   |                        |                           | <b>60,328</b>        | <b>63,935</b>        | <b>56,773</b>        | <b>181,036</b>        |

distribution of recombination and allelic diversity, can now be applied to wheat improvement strategies. This is especially true if “must-have traits” that are allocated in chromosomal compartments with highly contrasting characteristics are fixed in repulsion or are found only in incompatible gene pools of the respective breeding germplasm.

### Rapid trait improvement using physically resolved markers and genome editing

The selection and modification of genetic variation underlying agronomic traits in breeding programs is often complicated if phenotypic selection depends on the expression of multiple loci with quantitative effects that can be strongly influenced by the environment. This dilemma can be overcome if DNA markers in strong linkage disequilibrium with the phenotype are identified through forward genetic approaches or if the underlying genes can be targeted through genome editing. The potential for IWGSC RefSeq v1.0, together with the detailed genome annotation, to accelerate the identification of potential candidate genes underlying important agronomic traits was exemplified for two targets. A forward genetics approach was used to fully resolve a QTL for stem solidness (*SstI*) conferring resistance to drought stress and insect damage (49) that was disrupted in previous wheat assemblies by a lack of scaffold ordering and annotation, partial assembly, and/or incomplete gene models (fig. S57 and tables S40 and S41). In IWGSC RefSeq v1.0, *SstI* contains 160 HC genes (table S42), of which 26 were differentially expressed (DESeq2, Benjamini-Hochberg adjusted  $P < 0.01$ ) between wheat lines with contrasting phenotypes. One of the differentially expressed genes, *TraesCS3B01G608800*, was present as a single copy in IWGSC RefSeq v1.0 but showed copy number variation associated with stem solidness in a diverse panel of hexaploid cultivars (Fig. 5A, fig. S58, and table S43). Using IWGSC RefSeq v1.0, we developed a diagnostic SNP marker physically linked to the copy number variation that has been deployed to select for stem solidness in wheat breeding programs (Fig. 5B).

Knowledge from model species can also be used to annotate genes and provide a route to trait enhancement through reverse genetics. The approach here targeted flowering time, which is important for crop adaptation to diverse environments and is well studied in model plants. Six wheat homologs of the *FLOWERING LOCUS C* (*FLC*) gene have been identified as having a role in the vernalization response, a critical process regulating flowering time (50). IWGSC RefSeq v1.0 was used to refine the annotation of these six sequences to identify four HC genes and then to design guide RNAs to specifically target, with CRISPR-Cas9-based gene editing, one of these genes, *TaAGL33*, on all subgenomes [*TraesCS3A01G435000* (A), *TraesCS3B01G470000* (B), and *TraesCS3D01G428000* (D)] [Fig. 5C and (14)]. Editing was obtained at the targeted gene

and led to truncated proteins after the MADS box through small deletions and insertions (Fig. 5D). Expression of all homeologs was high before vernalization, dropped during vernalization, and remained low post-vernalization, implying a role for this gene in flowering control. This expression pattern was not strongly affected by the genome edits (fig. S59). Plants with the editing events in the D subgenome flowered 2 to 3 days earlier than controls (Fig. 5E). Further refinement should help to fully elucidate the importance of the *TaAGL33* gene for vernalization in monocots. These results exemplify how the IWGSC RefSeq v1.0 could accelerate the development of diagnostic markers and the design of targets for genome editing for traits relevant to breeding.

### Conclusions

IWGSC RefSeq v1.0 is a resource that has the potential for disruptive innovation in wheat improvement. By necessity, breeders work with the genome at the whole-chromosome level, as each new cross involves the modification of genome-wide gene networks that control the expression of complex traits such as yield. With the annotated and ordered reference genome sequence in place, researchers and breeders can now easily access sequence-level information to define changes in the genomes of lines in their programs. Although several hundred wheat QTLs have been published, only a small number of genes have been cloned and functionally characterized. IWGSC RefSeq v1.0 underpins immediate application by providing access to regulatory regions, and it will serve as the backbone to anchor all known QTLs to one common annotated reference. Combining this knowledge with the distribution of meiotic recombination frequency and genomic diversity will enable breeders to more efficiently tackle the challenges imposed by the need to balance the parallel selection processes for adaptation to biotic and abiotic stress, end-use quality, and yield improvement. Strategies can now be defined more precisely to bring desirable alleles into coupling phase, especially in less-recombinant regions of the wheat genome. Here the full potential of the newly available genome information may be realized through the implementation of DNA-marker platforms and targeted breeding technologies, including genome editing (57).

### Methods summary

Whole-genome sequencing of cultivar Chinese Spring by short-read sequencing-by-synthesis provided the data for de novo genome assembly and scaffolding with the software package DenovoMAGIC2. The assembly was superscaffolded and anchored into 21 pseudomolecules with high-density genetic (POPSEQ) and physical (Hi-C and 21 chromosome-specific physical maps) mapping information and by integrating additional genomic resources. Validation of the assembly used independent genetic (de novo genotyping-by-sequencing maps) and physical mapping evidence (radiation hybrid maps, BioNano “optical

maps” for group 7 homeologous chromosomes). The genome assembly was annotated for genes, repetitive DNA, and other genomic features, and in-depth comparative analyses were carried out to analyze the distribution of genes, recombination, position, and size of centromeres and the expansion and contraction of wheat gene families. An atlas of wheat gene transcription was built from an extensive panel of 850 independent transcriptome datasets and was then used to study gene coexpression networks. Furthermore, the assembly was used for the dissection of an important stem-solidness QTL and to design targets for genome editing of genes implicated in flowering-time control in wheat. Detailed methodological procedures are described in the supplementary materials.

### REFERENCES AND NOTES

- Food and Agriculture Organization of the United Nations, FAOSTAT statistics database, Food balance sheets (2017); [www.fao.org/faostat/en/#data/FBS](http://www.fao.org/faostat/en/#data/FBS).
- Food and Agriculture Organization of the United Nations, FAOSTAT statistics database, Crops (2017); [www.fao.org/faostat/en/#data/QC](http://www.fao.org/faostat/en/#data/QC).
- G. N. Atlin, J. E. Cairns, B. Das, Rapid breeding and varietal replacement are critical to adaptation of cropping systems in the developing world to climate change. *Glob. Food Sec.* **12**, 31–37 (2017). doi: [10.1016/j.gfs.2017.01.008](https://doi.org/10.1016/j.gfs.2017.01.008); pmid: [28580238](https://pubmed.ncbi.nlm.nih.gov/28580238/)
- J. M. Hickey, T. Chiurugwi, I. Mackay, W. Powell; Implementing Genomic Selection in CGIAR Breeding Programs Workshop Participants, Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat. Genet.* **49**, 1297–1303 (2017). doi: [10.1038/ng.3920](https://doi.org/10.1038/ng.3920); pmid: [28854179](https://pubmed.ncbi.nlm.nih.gov/28854179/)
- K. Arumuganathan, E. D. Earle, Nuclear DNA content of some important plant species. *Plant Mol. Biol. Report.* **9**, 208–218 (1991). doi: [10.1007/BF02672069](https://doi.org/10.1007/BF02672069)
- International Wheat Genome Sequencing Consortium (IWGSC), A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* **345**, 1251788 (2014). doi: [10.1126/science.1251788](https://doi.org/10.1126/science.1251788); pmid: [25035500](https://pubmed.ncbi.nlm.nih.gov/25035500/)
- J. A. Chapman *et al.*, A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biol.* **16**, 26 (2015). doi: [10.1186/s13059-015-0582-8](https://doi.org/10.1186/s13059-015-0582-8); pmid: [25637298](https://pubmed.ncbi.nlm.nih.gov/25637298/)
- B. J. Clavijo *et al.*, An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. *Genome Res.* **27**, 885–896 (2017). doi: [10.1101/gr.217117.116](https://doi.org/10.1101/gr.217117.116); pmid: [28420692](https://pubmed.ncbi.nlm.nih.gov/28420692/)
- A. V. Zimin *et al.*, The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum*. *Gigascience* **6**, 1–7 (2017). doi: [10.1093/gigascience/gix097](https://doi.org/10.1093/gigascience/gix097); pmid: [29069494](https://pubmed.ncbi.nlm.nih.gov/29069494/)
- T. R. Endo, B. S. Gill, The deletion stocks of common wheat. *J. Hered.* **87**, 295–307 (1996). doi: [10.1093/oxfordjournals.jhered.a023003](https://doi.org/10.1093/oxfordjournals.jhered.a023003)
- M. E. Sorrells *et al.*, Comparative DNA sequence analysis of wheat and rice genomes. *Genome Res.* **13**, 1818–1827 (2003). pmid: [12902377](https://pubmed.ncbi.nlm.nih.gov/12902377/)
- K. Eversole, J. Rogers, B. Keller, R. Appels, C. Feuillet, in *Breeding, Quality Traits, Pests and Diseases*, vol. 1 of *Achieving Sustainable Cultivation of Wheat*, P. Langridge, Ed. (Burlingame-Dodds Science Publishing, 2017), chap. 2.
- E. Paux *et al.*, Insertion site-based polymorphism markers open new perspectives for genome saturation and marker-assisted selection in wheat. *Plant Biotechnol. J.* **8**, 196–210 (2010). doi: [10.1111/j.1467-7652.2009.00477.x](https://doi.org/10.1111/j.1467-7652.2009.00477.x); pmid: [20078842](https://pubmed.ncbi.nlm.nih.gov/20078842/)
- See supplementary materials.
- Arabidopsis Genome Initiative, Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000). doi: [10.1038/35048692](https://doi.org/10.1038/35048692); pmid: [11130711](https://pubmed.ncbi.nlm.nih.gov/11130711/)
- International Rice Genome Sequencing Project, The map-based sequence of the rice genome. *Nature* **436**, 793–800 (2005). doi: [10.1038/nature03895](https://doi.org/10.1038/nature03895); pmid: [16100779](https://pubmed.ncbi.nlm.nih.gov/16100779/)
- T. Wicker *et al.*, International Wheat Genome Sequencing Consortium, Impact of transposable elements on genome



- structure and evolution in wheat. *Genome Biol.* doi: 10.1186/s13059-018-1479-0 (2018).
18. F. Choulet *et al.*, Structural and functional partitioning of bread wheat chromosome 3B. *Science* **345**, 1249721 (2014). doi: 10.1126/science.1249721; pmid: 25035497
  19. X. Guo *et al.*, De novo centromere formation and centromeric sequence expansion in wheat and its wide hybrids. *PLOS Genet.* **12**, e1005997 (2016). doi: 10.1371/journal.pgen.1005997; pmid: 27110907
  20. K. Wang, Y. Wu, W. Zhang, R. K. Dawe, J. Jiang, Maize centromeres expand and adopt a uniform size in the genetic background of oat. *Genome Res.* **24**, 107–116 (2014). doi: 10.1101/gr.160887.113; pmid: 24100079
  21. Y. Jiao *et al.*, Improved maize reference genome with single-molecule technologies. *Nature* **546**, 524–527 (2017). pmid: 28605751
  22. H. Yan *et al.*, Intergenic locations of rice centromeric chromatin. *PLOS Biol.* **6**, e286 (2008). doi: 10.1371/journal.pbio.0060286; pmid: 19067486
  23. R. H. Ramírez-González *et al.*, The transcriptional landscape of polyploid wheat. *Science* **361**, eaar6089 (2018). doi: 10.1126/science.aar6089
  24. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015). doi: 10.1093/bioinformatics/btv351; pmid: 26059717
  25. M.-D. Rey *et al.*, Exploiting the ZIP4 homologue within the wheat *Ph1* locus has identified two lines exhibiting homologous crossover in wheat-wild relative hybrids. *Mol. Breed.* **37**, 95 (2017). doi: 10.1007/s11032-017-0700-2; pmid: 28781573
  26. F. Cheng *et al.*, Gene retention, fractionation and subgenome differences in polyploid plants. *Nat. Plants* **4**, 258–268 (2018). doi: 10.1038/s41477-018-0136-7; pmid: 29725103
  27. T. Marcussen *et al.*, Ancient hybridizations among the ancestral genomes of bread wheat. *Science* **345**, 1250092 (2014). pmid: 25035499
  28. Y. Van de Peer, S. Maere, A. Meyer, The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* **10**, 725–732 (2009). doi: 10.1038/nrg2600; pmid: 19652647
  29. P. S. Soltis, D. E. Soltis, Ancient WGD events as drivers of key innovations in angiosperms. *Curr. Opin. Plant Biol.* **30**, 159–165 (2016). doi: 10.1016/j.cpb.2016.03.015; pmid: 27064530
  30. M. Melé *et al.*, The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015). doi: 10.1126/science.1250355; pmid: 25954002
  31. S. C. Stelpflug *et al.*, An expanded maize gene expression atlas based on RNA sequencing and its use to explore root development. *Plant Genome* **9**, 0025 (2016). pmid: 27898762
  32. F. He *et al.*, Large-scale atlas of microarray data reveals the distinct expression landscape of different tissues in Arabidopsis. *Plant J.* **86**, 472–480 (2016). doi: 10.1111/tj.13175; pmid: 27015116
  33. X. Wang *et al.*, Comparative genomic analysis of C4 photosynthetic pathway evolution in grasses. *Genome Biol.* **10**, R68 (2009). doi: 10.1186/gb-2009-10-6-r68; pmid: 19549309
  34. L. Xia *et al.*, Rice Expression Database (RED): An integrated RNA-Seq-derived gene expression database for rice. *J. Genet. Genomics* **44**, 235–241 (2017). doi: 10.1016/j.jgg.2017.05.003; pmid: 28529082
  35. R. J. Schaefer, J.-M. Michno, C. L. Myers, Unraveling gene function in agricultural species using gene co-expression networks. *Biochim. Biophys. Acta* **1860**, 53–63 (2017). pmid: 27485388
  36. J. Beales, A. Turner, S. Griffiths, J. W. Snape, D. A. Laurie, A *Pseudo-Response Regulator* is misexpressed in the photoperiod insensitive *Ppd-D1a* mutant of wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* **115**, 721–733 (2007). doi: 10.1007/s00122-007-0603-4; pmid: 17634915
  37. L. Yan *et al.*, The wheat and barley vernalization gene *VRN3* is an orthologue of *FT*. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19581–19586 (2006). doi: 10.1073/pnas.0607142103; pmid: 17158798
  38. K. V. Krasileva *et al.*, Uncovering hidden variation in polyploid wheat. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E913–E921 (2017). doi: 10.1073/pnas.1619268114; pmid: 28096351
  39. S. Wang *et al.*, Cytological and transcriptomic analyses reveal important roles of *CLE19* in pollen exine formation. *Plant Physiol.* **175**, 1186–1202 (2017). doi: 10.1104/pp.17.00439; pmid: 28916592
  40. M. Pfeifer *et al.*, Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science* **345**, 1250091 (2014). doi: 10.1126/science.1250091; pmid: 25035498
  41. P. Borrill, N. Adamski, C. Uauy, Genomics as the key to unlocking the polyploid potential of wheat. *New Phytol.* **208**, 1008–1022 (2015). doi: 10.1111/nph.13533; pmid: 26108556
  42. F. A. Kondrashov, Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc. Biol. Sci.* **279**, 5048–5057 (2012). doi: 10.1098/rspb.2012.1108; pmid: 22977152
  43. P. H. Schiffer, J. Gravemeyer, M. Rauscher, T. Wiehe, Ultra large gene families: A matter of adaptation or genomic parasites? *Life* **6**, 32 (2016). doi: 10.3390/life603032; pmid: 27509525
  44. T. Sykes *et al.*, In silico identification of candidate genes for fertility restoration in cytoplasmic male sterile perennial ryegrass (*Lolium perenne* L.). *Genome Biol. Evol.* **9**, 351–362 (2017). pmid: 26951780
  45. J. Melonek, J. D. Stone, I. Small, Evolutionary plasticity of restorer-of-fertility-like proteins in rice. *Sci. Rep.* **6**, 35152 (2016). doi: 10.1038/srep35152; pmid: 27775031
  46. T. Würschum, C. F. H. Longin, V. Hahn, M. R. Tucker, W. L. Leiser, Copy number variations of *CBF* genes at the *Fr-A2* locus are essential components of winter hardiness in wheat. *Plant J.* **89**, 764–773 (2017). doi: 10.1111/tj.13424; pmid: 27859852
  47. A. Juhász *et al.*, Genome mapping of seed-borne allergens and immune-responsive proteins in wheat. *Sci. Adv.* **4**, eaar8602 (2018). doi: 10.1126/sciadv.aar8602
  48. M. Feldman, A. A. Levy, in *Alien Introgression in Wheat: Cytogenetics, Molecular Biology, and Genomics*, M. Molnár-Láng, C. Ceoloni, J. Doležal, Eds. (Springer, 2015), pp. 21–76.
  49. K. T. Nilsen *et al.*, High density mapping and haplotype analysis of the major stem-solidness locus *Sst1* in durum and common wheat. *PLOS ONE* **12**, e0175285 (2017). doi: 10.1371/journal.pone.0175285; pmid: 28399136
  50. N. Sharma *et al.*, A flowering locus C homolog is a vernalization-regulated repressor in *Brachypodium* and is cold regulated in wheat. *Plant Physiol.* **173**, 1301–1315 (2017). doi: 10.1104/pp.16.01161; pmid: 28034954
  51. H. Puchta, Applying CRISPR/Cas for genome engineering in plants: The best is yet to come. *Curr. Opin. Plant Biol.* **36**, 1–8 (2017). doi: 10.1016/j.cpb.2016.11.011; pmid: 27914284
  52. International Wheat Genome Sequencing Consortium, Gene family expansion and contraction in the genome of bread wheat cv. Chinese Spring. *eDAL* (2018). doi: 10.5447/IPK/2018/5
  53. H.-Q. Ling *et al.*, Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature* **496**, 87–90 (2013). doi: 10.1038/nature11997; pmid: 23535596
  54. J. Jia *et al.*, *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* **496**, 91–95 (2013). doi: 10.1038/nature12028; pmid: 23535592
  55. Y. Ishida, M. Tsunashima, Y. Hiei, T. Komari, in *Agrobacterium Protocols: Volume 1*, K. Wang, Ed. (Springer, 2015), pp. 189–198.

# ACKNOWLEDGMENTS

The IWGSC would like to thank the following individuals: M. Burrell and C. Bridson (Norwich Biosciences Institute) for computational support of RNA-seq data; I. Christie (Grainor AS) and H. Rudi (Norwegian University of Life Sciences) for assistance with chromosome 7B; R. P. Davey (Earlham Institute) for assistance with RNA-seq data; J. Deek (Tel Aviv University) for growing the source plants and DNA extraction used for whole-genome sequencing; Z. Dubská, E. Jahnová, M. Seifertová, R. Šperková, R. Tušková, and J. Weiserová (Institute of Experimental Botany, Olomouc) for assistance with flow cytometric chromosome sorting, BAC library construction, and estimation of genome size; S. Durand, V. Jamilloux, M. Lainé, and C. Michotey (URGI, INRA) for assistance with and access to the IWGSC sequence repository; A. Fiebig of the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) for submitting the Hi-C data; T. Florio for the design of the wheat schematic for the expression atlas and *Sst1* figure ([www.flozbox.com/Science\\_Illustrated](http://www.flozbox.com/Science_Illustrated)); C. Karunakaran and T. Bond of the Canadian Light Source for performing CT imaging; J. Kawai, N. Kondo, H. Sano, N. Suzuki, M. Tagami, and H. Tarui of RIKEN for assistance with deep sequencing of chromosome 6B; H. Fujisawa, Y. Katayose, K. Kurita, S. Mori, Y. Mukai, and H. Sasaki of the Institute of Crop Science, NARO for assistance with deep sequencing of chromosome 6B; T. Matsumoto of Tokyo University of Agriculture for assistance with deep sequencing of chromosome 6B; P. Lenoble and C. Orvain of Genoscope for assistance in the sequencing of chromosome 1B; A. J. Lukaszewski of the University

of California, Riverside, and B. Friebe and J. Raupp of Kansas State University for providing seeds of wheat teleostic lines for chromosome sorting; C. Maulis (<https://polytypo.design>; <https://propepper.net>) for design and graphics of the prolamin superfamily chromosome map; M. Seifertová and H. Tvardiková of the Institute of Experimental Botany for assistance with BAC DNA extraction and sequencing for chromosomes 3DS, 4A, and 7DS; and I. Willick and K. Tanino of the University of Saskatchewan for their assistance in sample preparation and the use of lab facilities.

**Funding:** The authors would like to thank the following for their financial support of research that enabled the completion of the IWGSC RefSeq v1.0 Project: Agence Nationale pour la Recherche (ANR), ANR-11-BSV5-0015-Ploid-Ploid Wheat-Unravelling bases of polyploidy success in wheat and ANR-16-TERC-0026-01-3DWHEAT; Agriculture and Agri-Food Canada National Wheat Improvement Program and the AgriFlex Program; Alberta Wheat Development Commission through the Canadian Applied Triticum Genomics (CTAG2); Australian Government, Department of Industry, Innovation, Climate Change, Science, Research, and Tertiary Education, Australia China Science and Research Fund Group Mission (Funding Agreement ACSRF00542); Australian Research Council Centre of Excellence in Plant Energy Biology (CE140100008); Australian Research Council Laureate Fellowship (FL140100179); Bayer CropScience; Biotechnology and Biological Sciences Research Council (BBSRC) 20:20 Wheat (project number BB/J00426X/1), Institute Strategic Programme grant (BB/J004669/1), Designing Future Wheat (DFW) Institute Strategic Programme (BB/P016855/1), the Wheat Genomics for Sustainable Agriculture (BB/J003557/1), and the Anniversary Future Leader Fellowship (BB/M014045/1); Canada First Research Excellence Fund through the Designing Crops for Global Food Security initiative at the University of Saskatchewan; Council for Agricultural Research and Economics, Italy, through CREA-Interomics; Department of Biotechnology, Ministry of Science and Technology, Government of India File No. F grant no. BT/IWGSC/03/TF/2008; DFG (SFB924) for support of KFXM; European Commission through the *TriticaceaeGenome* (FP7-212019); France Génomique (ANR-10-INBS-09) Genome Canada through the CTAG2 project; Genome Prairie through the CTAG2 project; German Academic Exchange Service (DAAD) PPP Australien 116; German Federal Ministry of Food and Agriculture grant 2819103915 WHEATSEQ; German Ministry of Education and Research grant 03IA536 de.NBI; Global Institute for Food Security Genomics and Bioinformatics fund; Gordon and Betty Moore Foundation grant GBMF4725 to Two Blades Foundation; Grain Research Development Corporation (GRDC) Australia; Graminor AS NFR project 199387, Expanding the technology base for Norwegian wheat breeding; Sequencing wheat chromosome 7B; Illumina; INRA, French National Institute for Agricultural Research; International Wheat Genome Sequencing Consortium and its sponsors; Israel Science Foundation grants 999/12, 1137/17, and 1824/12; Junta de Andalucía, Spain, project P12-AGR-0482; MINECO (Spanish Ministry of Economy, Industry, and Competitiveness) project BIO2011-15237-E; Ministry of Agriculture, Forestry, and Fisheries of Japan through Genomics for Agricultural Innovation, KGS-1003 and through Genomics-based Technology for Agricultural Improvement, NGB-1003; Ministry of Education and Science of Russian Federation project RFMEFI60414X0106 and project RFMEFI60414 X0107; Ministry of Education, Youth, and Sport of the Czech Republic award no. LO1204 (National Program of Sustainability I); Nissin Flour Milling, Inc.; National Research Council of Canada Wheat Flagship program; Norwegian University of Life Sciences (NMBU) NFR project 199387, Expanding the technology base for Norwegian wheat breeding, Sequencing wheat chromosome 7B; National Science Foundation, United States, award (FAIN) 1339389, GPF-PG: Genome Structure and Diversity of Wheat and Its Wild Relatives, award DBI-0701916, and award IIP-1338897; Russian Science Foundation project 14-14-00161; Saskatchewan Ministry of Agriculture through the CTAG2 project; Saskatchewan Wheat Development Commission through the CTAG2 project; The Czech Science Foundation award no. 521/06/1723 (Construction of BAC library and physical mapping of the wheat chromosome 3D), award no. 521-08-1629 (Construction of BAC DNA libraries specific for chromosome 4AL and positional cloning of gene for adult plant resistance to powdery mildew in wheat), award no. P501/10/1740 (Physical map of the wheat chromosome 4AL and positional cloning of a gene for yield), award no. P501/12/2554 (Physical map of wheat chromosome arm 7DS and its use to clone a Russian wheat aphid-resistance gene), award no. P501/12/G090 (Evolution and function of complex plant genomes), award no. 14-07164S (Cloning and molecular characterization of wheat *Qpm-tur-4A* gene conferring seedling and adult plant race-nonspecific powdery-mildew resistance), and award no. 13-08786S (Chromosome arm 3DS of bread wheat: Its

sequence and function in allopolyploid genome); The Research Council of Norway (NFR) project 199387, Expanding the technology base for Norwegian wheat breeding; Sequencing wheat chromosome 7B; U.S. Department of Agriculture NIFA 2008-35300-04588, the University of Zurich; Western Grains Research Foundation through the CTAG2 project; Western Grains Research Foundation National Wheat Improvement Program; and the Winifred-Asbjornson Plant Science Endowment Fund. The research leading to these results has also received funding from the French Government managed by the ANR under the Investment for the Future program (BreedWheat project ANR-10-BTBR-03), from FranceAgrisilv (2011-0971 and 2013-0544), French Funds to support Plant Breeding (FSOV), and INRA. Axiom genotyping was conducted on the genotyping platform GENTYANE at INRA Clermont-Ferrand ([gentyane.clermont.inra.fr](http://gentyane.clermont.inra.fr)). This research was supported in part by the NBI Computing infrastructure for Science (CIS) group through the HPC cluster.

**Author contributions:** See below, where authors are arranged by working group and contributions; leaders, co-leaders, and major contributors are listed alphabetically first and then other contributors follow alphabetically. **Competing interests:** The authors declare no competing interests. Bayer CropScience holds a patent application (WO2015000914A1) that covers the modulation of flowering time in monocots using the *FLC* gene. **Data and materials availability:** The IWGSC RefSeq v1.0 assembly and annotation data, physical maps for all chromosomes and chromosome arms, and all data related to this study are available in the IWGSC Data Repository hosted at URGI: <https://wheat-urgi.versailles.inra.fr/Seq-Repository>. Assembly and annotation data are also available at ENSEMBL-Plants: [https://plants.ensembl.org/Triticum\\_aestivum/Info/Index](https://plants.ensembl.org/Triticum_aestivum/Info/Index). The BAC libraries for all chromosomes and chromosome arms are available at the CNRGV-INRA: <http://cnrgv.toulouse.inra.fr/en/Library/Wheat>. Details on genome physical expansion and contraction in the genome of bread wheat cultivar Chinese Spring are provided in database S6 at <http://dx.doi.org/10.5447/IPK/2018/5> (52). The raw sequencing data used for de novo whole-genome assembly are available from the Sequence Read Archive under accession number SRP114784. RNA-seq data are available at SRA under accession numbers PRJEB25639, PRJEB23056, PRJNA436817, PRJEB25640, SRP133837, and PRJEB25593. Hi-C sequence data are available under accession number PRJEB25248. ChIP-seq data are available under SRA study PRJNA420988 (SRP1262229). CS bisulfite sequencing data are available under project number SRP133674 (SRG6792673 to SRR6792689). Organellar DNA sequences were deposited at NCBI GenBank (MH051715 and MH051716). Further details on data accessibility are outlined in the supplementary materials and methods.

## The International Wheat Genome Sequencing Consortium (IWGSC)

**IWGSC RefSeq principal investigators:** Rudi Appels<sup>1,36\*</sup>, Kellye Eversole<sup>2,3\*</sup>, Catherine Feuillet<sup>17</sup>, Beat Keller<sup>41</sup>, Jane Rogers<sup>6\*</sup>, Nils Stein<sup>4,5\*</sup>.

**IWGSC whole-genome assembly principal investigators:** Curtis J. Pozniak<sup>11</sup>, Nils Stein<sup>4,5\*</sup>, Frédéric Choulet<sup>7</sup>, Assaf Distelfeld<sup>25</sup>, Kellye Eversole<sup>2,3\*</sup>, Jesse Poland<sup>28</sup>, Jane Rogers<sup>6</sup>, Gil Ronen<sup>12</sup>, Andrew G. Sharpe<sup>43</sup>.

**Whole-genome sequencing and assembly:** Curtis Pozniak<sup>11</sup>, Gil Ronen<sup>12</sup>, Nils Stein<sup>4,5\*</sup>, Omer Barad<sup>17</sup>, Kobi Baruch<sup>12</sup>, Frédéric Choulet<sup>7</sup>, Gabriel Keeble-Gagnère<sup>1</sup>, Martin Mascher<sup>4,67</sup>, Andrew G. Sharpe<sup>43</sup>, Gil Ben-Zvi<sup>12</sup>, Ambre-Aurora Josselin<sup>7</sup>. **Hi-C data-based scaffolding:** Nils Stein<sup>4,5\*</sup>, Martin Mascher<sup>4,67</sup>, Axel Himmelbach<sup>4</sup>.

**Whole-genome assembly quality control and analyses:** Frédéric Choulet<sup>7</sup>, Gabriel Keeble-Gagnère<sup>1</sup>, Martin Mascher<sup>4,67</sup>, Jane Rogers<sup>6\*</sup>, François Balfourier<sup>7</sup>, Juan Gutierrez-Gonzalez<sup>30</sup>, Matthew Hayden<sup>2</sup>, Ambre-Aurora Josselin<sup>7</sup>, ChuShin Koh<sup>43</sup>, Gary Muehlbauer<sup>30</sup>, Raj K. Pasam<sup>1</sup>, Etienne Paux<sup>7</sup>, Curtis J. Pozniak<sup>11</sup>, Philippe Rigault<sup>39</sup>, Andrew G. Sharpe<sup>43</sup>, Josquin Tibbits<sup>15</sup>, Vijay Tiwari<sup>54</sup>. **Pseudomolecule assembly:** Frédéric Choulet<sup>7</sup>, Gabriel Keeble-Gagnère<sup>1</sup>, Martin Mascher<sup>4,67</sup>, Ambre-Aurora Josselin<sup>7</sup>, Jane Rogers<sup>6</sup>.

**RefSeq genome structure and gene analyses:** Manuel Spannagel<sup>9</sup>, Frédéric Choulet<sup>7</sup>, Daniel Lang<sup>9</sup>, Heidrun Gundlach<sup>9</sup>, Georg Haberer<sup>2</sup>, Gabriel Keeble-Gagnère<sup>1</sup>, Klaus F. X. Mayer<sup>9,44</sup>, Danara Ormanbekova<sup>9,48</sup>, Etienne Paux<sup>7</sup>, Verena Prade<sup>9</sup>, Hana Šimková<sup>8</sup>, Thomas Wicker<sup>41</sup>. **Automated annotation:** Frédéric Choulet<sup>7</sup>, Manuel Spannagel<sup>9</sup>, David Swarbreck<sup>50</sup>, Hélène Rimbart<sup>7</sup>, Marius Felder<sup>9</sup>, Nicolas Guillhot<sup>7</sup>, Heidrun Gundlach<sup>9</sup>, Georg Haberer<sup>2</sup>, Gemy Kaithakkottai<sup>50</sup>, Jens Kellwagen<sup>40</sup>, Daniel Lang<sup>9</sup>, Philippe Leroy<sup>7</sup>, Thomas Lux<sup>9</sup>, Klaus F. X. Mayer<sup>9,44</sup>, Sven Twardziok<sup>9</sup>, Luca Venturini<sup>50</sup>.

**Manual gene curation:** Rudi Appels<sup>1,36\*</sup>, Hélène Rimbart<sup>7</sup>, Frédéric Choulet<sup>7</sup>, Angéla Juhász<sup>36,37</sup>, Gabriel Keeble-Gagnère<sup>1</sup>.

**Subgenome comparative analyses:** Frédéric Choulet<sup>7</sup>, Manuel Spannagel<sup>9</sup>, Daniel Lang<sup>9</sup>, Michael Abrouk<sup>8,19</sup>, Georg Haberer<sup>2</sup>, Gabriel Keeble-Gagnère<sup>1</sup>, Klaus F. X. Mayer<sup>9,44</sup>, Thomas Wicker<sup>41</sup>.

**Transposable elements:** Frédéric Choulet<sup>7</sup>, Thomas Wicker<sup>41</sup>, Heidrun Gundlach<sup>9</sup>, Daniel Lang<sup>9</sup>, Manuel Spannagel<sup>9</sup>.

**Phylogenomic analyses:** Daniel Lang<sup>9</sup>, Manuel Spannagel<sup>9</sup>, Rudi Appels<sup>1,36\*</sup>, Iris Fischer<sup>9</sup>.

**Transcriptome analyses and RNA-seq data:** Cristobal Uauy<sup>10</sup>, Philippa Borriell<sup>10</sup>, Ricardo H. Ramirez-Gonzalez<sup>10</sup>.

Rudi Appels<sup>1,36\*</sup>, Dominique Arnaud<sup>63</sup>, Smahane Chalabi<sup>63</sup>, Boulos Chalhoub<sup>62,63</sup>, Frédéric Choulet<sup>7</sup>, Aron Cory<sup>11</sup>, Raju Datla<sup>22</sup>, Mark W. Davey<sup>18</sup>, Matthew Hayden<sup>2</sup>, John Jacobs<sup>18</sup>, Daniel Lang<sup>9</sup>, Stephen J. Robinson<sup>52</sup>, Manuel Spannagel<sup>9</sup>, Burkhard Steuernagel<sup>10</sup>, Josquin Tibbits<sup>15</sup>, Vijay Tiwari<sup>54</sup>, Fred van Ex<sup>18</sup>, Brande B. H. Wulff<sup>10</sup>.

**Whole-genome methylation:** Curtis J. Pozniak<sup>11</sup>, Stephen J. Robinson<sup>52</sup>, Andrew G. Sharpe<sup>43</sup>, Aron Cory<sup>11</sup>.

**Histone mark analyses:** Moussa Benhamed<sup>12</sup>, Etienne Paux<sup>7</sup>, Abdelhadi Bendahmane<sup>15</sup>, Lorenzo Concia<sup>15</sup>, David Latrasse<sup>15</sup>.

**BAC chromosome MTP IWGSC-Bayer Whole-Genome Profiling (WGP) tags:** Jane Rogers<sup>6\*</sup>, John Jacobs<sup>18</sup>, Michael Alaux<sup>13</sup>, Rudi Appels<sup>1,36\*</sup>, Jan Bartoš<sup>8</sup>, Arnaud Bellec<sup>20</sup>, Hélène Berges<sup>20</sup>, Jaroslav Doležel<sup>8</sup>, Catherine Feuillet<sup>17</sup>, Zeev Frenkel<sup>26</sup>, Bikram Gill<sup>28</sup>, Abraham Korol<sup>26</sup>, Thomas Letellier<sup>13</sup>, Odd-Arne Olsen<sup>56</sup>, Hana Šimková<sup>8</sup>, Kuldeep Singh<sup>65</sup>, Miroslav Valárik<sup>8</sup>, Edwin van der Vossen<sup>64</sup>, Sonia Vautrin<sup>20</sup>, Song Weining<sup>66</sup>.

**Chromosome LTC mapping and physical mapping quality control:** Abraham Korol<sup>26</sup>, Zeev Frenkel<sup>26</sup>, Tzion Fahima<sup>26</sup>, Vladimir Glikson<sup>29</sup>, Dina Raats<sup>50</sup>, Jane Rogers<sup>6</sup>.

**RH mapping:** Vijay Tiwari<sup>54</sup>, Bikram Gill<sup>28</sup>, Etienne Paux<sup>7</sup>, Jesse Poland<sup>28</sup>.

**Optical mapping:** Jaroslav Doležel<sup>8</sup>, Jarmila Čiháliková<sup>8</sup>, Hana Šimková<sup>8</sup>, Helena Toegolova<sup>8</sup>, Jan Vrána<sup>8</sup>.

**Recombination analyses:** Pierre Sourdil<sup>17</sup>, Benoit Darrier<sup>7</sup>.

**Gene family analyses:** Rudi Appels<sup>1,36\*</sup>, Manuel Spannagel<sup>9</sup>, Daniel Lang<sup>9</sup>, Iris Fischer<sup>9</sup>, Danara Ormanbekova<sup>9,48</sup>, Verena Prade<sup>9</sup>.

**CBF gene family:** Delfina Barabasi<sup>15</sup>, Luigi Cattivelli<sup>16</sup>.

**Dehydrin gene family:** Pilar Hernandez<sup>23</sup>, Sergio Galvez<sup>27</sup>, Hikmet Budak<sup>14</sup>.

**NLR gene family:** Burkhard Steuernagel<sup>10</sup>, Jonathan D. G. Jones<sup>35</sup>, Kamil Witek<sup>35</sup>, Brande B. H. Wulff<sup>10</sup>, Guotai Yu<sup>10</sup>.

**PPR gene family:** Ian Small<sup>45</sup>, Joanna Melonek<sup>45</sup>, Ruonan Zhou<sup>7</sup>.

**Prolamin gene family:** Angéla Juhász<sup>36,37</sup>, Tatiana Belova<sup>56</sup>, Rudi Appels<sup>1,36\*</sup>, Odd-Arne Olsen<sup>56</sup>.

**WAK gene family:** Kostya Kanyuka<sup>38</sup>, Robert King<sup>42</sup>.

**Stem solidness (SST1) QTL team:** Kirby Nilsen<sup>11</sup>, Sean Walkowiak<sup>11</sup>, Curtis J. Pozniak<sup>11</sup>, Richard Cuthbert<sup>21</sup>, Raju Datla<sup>22</sup>, Ron Knox<sup>41</sup>, Krysta Wiebe<sup>11</sup>, Daoquan Xiang<sup>22</sup>.

**Flowering locus C (FLC) gene team:** Antje Rohde<sup>72</sup>, Timothy Golds<sup>18</sup>.

**Genome size analysis:** Jaroslav Doležel<sup>8</sup>, Jana Čížková<sup>8</sup>, Josquin Tibbits<sup>15</sup>.

**MicroRNA and tRNA annotation:** Hikmet Budak<sup>14</sup>, Bala Ani Akpinar<sup>14</sup>, Sezgi Biyikoglu<sup>14</sup>.

**Genetic maps and mapping:** Gary Muehlbauer<sup>30</sup>, Jesse Poland<sup>28</sup>, Liangliang Gao<sup>28</sup>, Juan Gutierrez-Gonzalez<sup>30</sup>, Amidou N'Daïye<sup>11</sup>.

**BAC libraries and chromosome sorting:** Jaroslav Doležel<sup>8</sup>, Hana Šimková<sup>8</sup>, Jarmila Čiháliková<sup>8</sup>, Marie Kubaláková<sup>8</sup>, Jan Šatfár<sup>8</sup>, Jan Vrána<sup>8</sup>.

**BAC pooling, BAC library repository, and access:** Hélène Berges<sup>20</sup>, Arnaud Bellec<sup>20</sup>, Sonia Vautrin<sup>20</sup>.

**IWGSC sequence and data repository and access:** Michael Alaux<sup>13</sup>, Françoise Alfama<sup>13</sup>, Anne-Françoise Adam-Blondon<sup>13</sup>, Raphaël Flores<sup>13</sup>, Claire Guerche<sup>13</sup>, Thomas Letellier<sup>13</sup>, Mikael Loaec<sup>13</sup>, Hadi Quesneville<sup>13</sup>.

**Physical maps and BAC-based sequences:**

**1A BAC sequencing and assembly:** Curtis J. Pozniak<sup>11</sup>, Andrew G. Sharpe<sup>42,43</sup>, Sean Walkowiak<sup>11</sup>, Hikmet Budak<sup>14</sup>, Janet Condie<sup>22</sup>, Jennifer Ens<sup>11</sup>, ChuShin Koh<sup>43</sup>, Ron MacLachlan<sup>11</sup>, Yifang Tan<sup>22</sup>, Thomas Wicker<sup>41</sup>.

**1B BAC sequencing and assembly:** Frédéric Choulet<sup>7</sup>, Etienne Paux<sup>7</sup>, Adriana Alberti<sup>61</sup>, Jean-Marc Aury<sup>61</sup>, François Balfourier<sup>7</sup>, Valérie Barbe<sup>61</sup>, Arnaud Couloux<sup>61</sup>, Corinne Cruaud<sup>61</sup>, Karine Labadie<sup>61</sup>, Sophie Mangenot<sup>61</sup>, Patrick Wincker<sup>61,68,69</sup>.

**1D, 4D, and 6D physical mapping:** Bikram Gill<sup>28</sup>, Gaganpreet Kaur<sup>28</sup>, Mingcheng Luo<sup>34</sup>, Sunish Sehgal<sup>53</sup>.

**2AL physical mapping:** Kuldeep Singh<sup>65</sup>, Parveen Chhuneja<sup>65</sup>, Om Prakash Gupta<sup>65</sup>, Suruchi Jindal<sup>65</sup>, Parampreet Kaur<sup>65</sup>, Palvi Malik<sup>65</sup>, Priti Sharma<sup>65</sup>, Bharat Yadav<sup>65</sup>.

**2AS physical mapping:** Nagendra K. Singh<sup>70</sup>, Jitendra P. Khurana<sup>71</sup>, Chandanker Chaudhary<sup>71</sup>, Paramjit Khurana<sup>71</sup>, Vinod Kumar<sup>70</sup>, Ajay Mahato<sup>70</sup>, Saloni Mathur<sup>71</sup>, Amitha Sevanthi<sup>70</sup>, Naveen Sharma<sup>71</sup>, Ram Sewak Tomar<sup>70</sup>.

**2B, 2D, 4B, 5BL, and 5DL IWGSC-Bayer Whole-Genome**

**Profiling (WGP) physical maps:** Jane Rogers<sup>6\*</sup>, John Jacobs<sup>18</sup>, Michael Alaux<sup>13</sup>, Arnaud Bellec<sup>20</sup>, Hélène Berges<sup>20</sup>, Jaroslav Doležel<sup>8</sup>, Catherine Feuillet<sup>17</sup>, Zeev Frenkel<sup>26</sup>, Bikram Gill<sup>28</sup>, Abraham Korol<sup>26</sup>, Edwin van der Vossen<sup>64</sup>, Sonia Vautrin<sup>20</sup>.

**3AL physical mapping:** Bikram Gill<sup>28</sup>, Gaganpreet Kaur<sup>28</sup>, Mingcheng Luo<sup>34</sup>, Sunish Sehgal<sup>53</sup>.

**3DS physical mapping and BAC sequencing and assembly:**

Jan Bartoš<sup>8</sup>, Kateřina Holušová<sup>8</sup>, Ondřej Plihal<sup>49</sup>.

**3DL BAC sequencing and assembly:** Matthew D. Clark<sup>50,73</sup>, Darren Heavens<sup>50</sup>, George Kettleborough<sup>50</sup>, Jon Wright<sup>50</sup>.

**4A physical mapping, BAC sequencing, assembly, and annotation:**

Miroslav Valárik<sup>8</sup>, Michael Abrouk<sup>8,19</sup>, Barbora Balčárková<sup>8</sup>, Kateřina Holušová<sup>8</sup>, Yuguin Hu<sup>34</sup>, Mingcheng Luo<sup>34</sup>.

**5BS BAC sequencing and assembly:** Elena Salina<sup>47</sup>.

Nikolai Ravin<sup>23,51</sup>, Konstantin Skryabin<sup>23,51</sup>, Alexey Beletsky<sup>23</sup>, Vitaly Kadnikov<sup>23</sup>, Andrey Mardanov<sup>23</sup>, Mikhail Nesterov<sup>47</sup>, Andrey Rakitin<sup>23</sup>, Ekaterina Sergeeva<sup>47</sup>.

**6B BAC sequencing and assembly:** Hirokazu Handa<sup>31</sup>.

Hiroyuki Kanamori<sup>31</sup>, Satoshi Katagiri<sup>31</sup>, Fuminori Kobayashi<sup>31</sup>, Shuhei Nasuda<sup>46</sup>, Tsuyoshi Tanaka<sup>31</sup>, Jianzhong Wu<sup>31</sup>.

**7A physical mapping and BAC sequencing:** Rudi Appels<sup>1,36\*</sup>, Matthew Hayden<sup>2</sup>, Gabriel Keeble-Gagnère<sup>1</sup>, Philippe Rigault<sup>39</sup>, Josquin Tibbits<sup>15</sup>.

**7B physical mapping, BAC sequencing, and assembly:**

Odd-Arne Olsen<sup>56</sup>, Tatiana Belova<sup>56</sup>, Federica Cattaroni<sup>58</sup>, Min Jiumeng<sup>60</sup>, Karl Kugler<sup>9</sup>, Klaus F. X. Mayer<sup>9,44</sup>, Matthias Pfeifer<sup>9</sup>, Simen Sandve<sup>57</sup>, Xu Xun<sup>59</sup>, Bujie Zhan<sup>56</sup>.

**7DS BAC sequencing and assembly:** Hana Šimková<sup>8</sup>.

Michael Abrouk<sup>8,19</sup>, Jacqueline Batley<sup>24</sup>, Philipp E. Bayer<sup>24</sup>, David Edwards<sup>24</sup>, Satomi Hayashi<sup>32</sup>, Helena Toegolova<sup>8</sup>, Zuzana Tulpová<sup>24</sup>, Paul Visendi<sup>25</sup>.

**7DL physical mapping and BAC sequencing:** Song Weining<sup>66</sup>, Licao Cui<sup>66</sup>, Xianghong Du<sup>66</sup>, Kewei Feng<sup>66</sup>, Xiaojun Nie<sup>66</sup>, Wei Tong<sup>66</sup>, Le Wang<sup>66</sup>.

**Figures:** Philippa Borriell<sup>10</sup>, Heidrun Gundlach<sup>9</sup>, Sergio Galvez<sup>27</sup>, Gemy Kaithakkottai<sup>50</sup>, Daniel Lang<sup>9</sup>, Thomas Lux<sup>9</sup>.

Martin Mascher<sup>4,67</sup>, Danara Ormanbekova<sup>9,48</sup>, Verena Prade<sup>9</sup>, Ricardo H. Ramirez-Gonzalez<sup>10</sup>, Manuel Spannagel<sup>9</sup>, Nils Stein<sup>4,5\*</sup>, Cristobal Uauy<sup>10</sup>, Luca Venturini<sup>50</sup>.

**Manuscript writing team:** Nils Stein<sup>4,5\*</sup>, Rudi Appels<sup>1,36\*</sup>.

Kellye Eversole<sup>2,3\*</sup>, Jane Rogers<sup>6\*</sup>, Philippa Borriell<sup>10</sup>, Luigi Cattivelli<sup>16</sup>, Frédéric Choulet<sup>7</sup>, Pilar Hernandez<sup>23</sup>.

Kostya Kanyuka<sup>38</sup>, Daniel Lang<sup>9</sup>, Martin Mascher<sup>4,67</sup>, Kirby Nilsen<sup>11</sup>, Etienne Paux<sup>7</sup>, Curtis J. Pozniak<sup>11</sup>, Ricardo H. Ramirez-Gonzalez<sup>10</sup>, Hana Šimková<sup>8</sup>, Ian Small<sup>45</sup>, Manuel Spannagel<sup>9</sup>, David Swarbreck<sup>50</sup>, Cristobal Uauy<sup>10</sup>.

<sup>1</sup>AgriBio, Centre for AgriBioscience, Department of Economic Development, Jobs, Transport, and Resources, 5 Ring Road, La Trobe University, Bundoora, VIC 3083, Australia. <sup>2</sup>International Wheat Genome Sequencing Consortium (IWGSC), 5207 Wyoming Road, Bethesda, MD 20816, USA. <sup>3</sup>Eversole Associates, 5207 Wyoming Road, Bethesda, MD 20816, USA. <sup>4</sup>Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Genebank, Corrensstr. 3, 06466 Stadt Seeland, Germany. <sup>5</sup>The University of Western Australia (UWA), School of Agriculture and Environment, 35 Stirling Highway, Crawley, WA 6009, Australia. <sup>6</sup>International Wheat Genome Sequencing Consortium (IWGSC), 18 High Street, Little Eversden, Cambridge CB23 1HE, UK. <sup>7</sup>GDEC (Genetics, Diversity and Ecophysiology of Cereals), INRA, Université Clermont Auvergne (UCA), 5 chemin de Beaulieu, 63039 Clermont-Ferrand, France. <sup>8</sup>Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Šlechtitelská 31, CZ-78371 Olomouc, Czech Republic. <sup>9</sup>Helmholtz Center Munich, Plant Genome and Systems Biology (PGSB), Ingolstaedter Landstr. 1, 85764 Neuherberg, Germany. <sup>10</sup>John Innes Centre, Crop Genetics, Norwich Research Park, Norwich NR4 7UH, UK. <sup>11</sup>University of Saskatchewan, Crop Development Centre, Agriculture Building, 51 Campus Drive, Saskatoon, SK S7N 5A8, Canada. <sup>12</sup>NRGene Ltd., 5 Golda Meir Street, Ness Ziona 7403648, Israel. <sup>13</sup>URGI, INRA, Université Paris-Saclay, 78026 Versailles, France. <sup>14</sup>Plant Sciences and Plant Pathology, Cereal Genomics Lab, Montana State University, 412 Leon Johnson Hall, Bozeman, MT 59717, USA. <sup>15</sup>Biology Department, Institute of Plant Sciences-Paris-Saclay, Bâtiment 630, rue de Noetlinz, Plateau du Moulon, CS80004, 91192 Gif-sur-Yvette Cedex, France. <sup>16</sup>Council for Agricultural Research and Economics (CREA), Research Centre for Genomics and Bioinformatics, via S. Protaso, 302, I -29017



Fiorenzuola d'Arda, Italy. <sup>17</sup>Bayer CropScience, Crop Science Division, Research and Development, Innovation Centre, 3500 Paramount Parkway, Morrisville, NC 27560, USA. <sup>18</sup>Bayer CropScience, Trait Research, Innovation Center, Technologiepark 38, 9052 Gent, Belgium. <sup>19</sup>Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia. <sup>20</sup>INRA, CNRGV, chemin de Borde Rouge, CS 52627, 31326 Castanet-Tolosan Cedex, France. <sup>21</sup>Agriculture and Agri-Food Canada, Swift Current Research and Development Centre, Box 1030, Swift Current, SK S9H 3X2, Canada. <sup>22</sup>National Research Council Canada, Aquatic and Crop Resource Development, 110 Gymnasium Place, Saskatoon, SK S7N 0W9, Canada. <sup>23</sup>Research Center of Biotechnology of the Russian Academy of Sciences, Institute of Bioengineering, Leninsky Avenue 33, Building 2, Moscow 119071, Russia. <sup>24</sup>School of Biological Sciences and Institute of Agriculture, University of Western Australia, Perth, WA 6009, Australia. <sup>25</sup>School of Plant Sciences and Food Security, Tel Aviv University, Ramat Aviv 69978, Israel. <sup>26</sup>University of Haifa, Institute of Evolution and the Department of Evolutionary and Environmental Biology, 199 Abba-Hushi Avenue, Mount Carmel, Haifa 3498838, Israel. <sup>27</sup>Universidad de Málaga. Lenguajes y Ciencias de la Computación, Campus de Teatinos, 29071 Málaga, Spain. <sup>28</sup>Plant Pathology, Throckmorton Hall, Kansas State University, Manhattan, KS 66506, USA. <sup>29</sup>MultiQTL Ltd., University of Haifa, Haifa 3498838, Israel. <sup>30</sup>Department of Agronomy and Plant Genetics, University of Minnesota, 411 Borlaug Hall, St. Paul, MN 55108, USA. <sup>31</sup>Institute of Crop Science, NARO, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8518, Japan. <sup>32</sup>Queensland University of Technology, Earth, Environmental and Biological Sciences, Brisbane, QLD 4001, Australia. <sup>33</sup>Instituto de Agricultura Sostenible (IAS-CSIC), Consejo Superior de Investigaciones Científicas, Alameda del Obispo s/n, 14004 Córdoba, Spain. <sup>34</sup>Department of Plant Sciences, University of California, Davis, One Shield Avenue, Davis, CA 95617, USA. <sup>35</sup>The Sainsbury Laboratory, Norwich Research Park, Norwich NR4 7UH, UK. <sup>36</sup>Murdoch University, Australia-China Centre for Wheat Improvement, School of Veterinary and Life Sciences, 90 South Street, Murdoch, WA 6150, Australia. <sup>37</sup>Agricultural Institute, MTA Centre for Agricultural Research, Applied Genomics Department, 2 Brunszvik Street, Martonvásár H 2462, Hungary. <sup>38</sup>Rothamsted Research,

Biointeractions and Crop Protection, West Common, Harpenden AL5 2JQ, UK. <sup>39</sup>GYDLE, Suite 220, 1135 Grande Allée, Ouest, Québec, QC G1S 1E7, Canada. <sup>40</sup>Julius Kühn-Institut, Institute for Biosafety in Plant Biotechnology, Erwin-Baur-Str. 27, 06484 Quedlinburg, Germany. <sup>41</sup>Department of Plant and Microbial Biology, University of Zurich, Zollikerstrasse 107, 8008 Zurich, Switzerland. <sup>42</sup>Rothamsted Research, Computational and Analytical Sciences, West Common, Harpenden AL5 2JQ, UK. <sup>43</sup>University of Saskatchewan, Global Institute for Food Security, 110 Gymnasium Place, Saskatoon, SK S7N 4J8, Canada. <sup>44</sup>School of Life Sciences Weihenstephan, Technical University of Munich, 85354 Freising, Germany. <sup>45</sup>School of Molecular Sciences, ARC Centre of Excellence in Plant Energy Biology, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia. <sup>46</sup>Graduate School of Agriculture, Kyoto University, Kitashirakawa-iwake-cho, Sakyo-ku, Kyoto 606-8502, Japan. <sup>47</sup>The Federal Research Center Institute of Cytology and Genetics, SB RAS, pr. Lavrentyeva 10, Novosibirsk 630090, Russia. <sup>48</sup>Department of Agricultural Sciences, University of Bologna, Viale Fanin, 44 40127 Bologna, Italy. <sup>49</sup>Department of Molecular Biology, Centre of the Region Haná for Biotechnological and Agricultural Research, Palacký University, Šlechtitelů 27, CZ-78371 Olomouc, Czech Republic. <sup>50</sup>Earlham Institute, Core Bioinformatics, Norwich NR4 7UZ, UK. <sup>51</sup>Faculty of Biology, Moscow State University, Leninskie Gory, 1, Moscow 119991, Russia. <sup>52</sup>Agriculture and Agri-Food Canada, Saskatoon Research and Development Centre, 107 Science Place, Saskatoon, SK S7N 0X2, Canada. <sup>53</sup>Agronomy Horticulture and Plant Science, South Dakota State University, 2108 Jackrabbit Drive, Brookings, SD 57006, USA. <sup>54</sup>Plant Science and Landscape Architecture, University of Maryland, 4291 Fieldhouse Road, 2102 Plant Sciences Building, College Park, MD 20742, USA. <sup>55</sup>University of Greenwich, Natural Resources Institute, Central Avenue, Chatham, Kent ME4 4TB, UK. <sup>56</sup>Faculty of Bioscience, Department of Plant Science, Norwegian University of Life Sciences, Arboretveien 6, 1433 Ås, Norway. <sup>57</sup>Faculty of Bioscience, Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Arboretveien 6, 1433 Ås, Norway. <sup>58</sup>Instituto di Genomica Applicata, Via J. Linussio 51, Udine 33100, Italy. <sup>59</sup>BGI-Shenzhen, BGI Genomics, Yantian District, Shenzhen 518083, Guangdong, China. <sup>60</sup>BGI-Shenzhen, BGI Genomics, Building No. 7, BGI Park, No. 21 Hongan 3rd Street,

Yantian District, Shenzhen 518083, China. <sup>61</sup>CEA-Institut de Biologie François-Jacob, Genoscope, 2 rue Gaston Crémieux, 91057 Evry Cedex, France. <sup>62</sup>Monsanto SAS, 28000 Boissay, France. <sup>63</sup>Institut National de la Recherche Agronomique (INRA), 2 rue Gaston Crémieux, 91057 Evry Cedex, France. <sup>64</sup>Keygene, N.V., Agro Business Park 90, 6708 PW Wageningen, Netherlands. <sup>65</sup>Punjab Agricultural University, Ludhiana, School of Agricultural Biotechnology, ICAR-National Bureau of Plant Genetic Resources, Dev Prakash Shastri Marg, New Delhi 110012, India. <sup>66</sup>State Key Laboratory of Crop Stress Biology in Arid Areas, College of Agronomy, Northwest A&F University, Yangling, Shaanxi 712101, China. <sup>67</sup>German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany. <sup>68</sup>CNRS, UMR 8030, CP5706, 91057 Evry, France. <sup>69</sup>Université d'Evry, UMR 8030, CP5706, 91057 Evry, France. <sup>70</sup>ICAR-National Research Centre on Plant Biotechnology, LBS Building, Pusa Campus, New Delhi 110012, India. <sup>71</sup>University of Delhi South Campus, Interdisciplinary Center for Plant Genomics and Department of Plant Molecular Biology, Benito Juarez Road, New Delhi 110021, India. <sup>72</sup>Bayer CropScience, Breeding and Trait Development, Technologiepark 38, 9052 Gent, Belgium. <sup>73</sup>Department of Lifesciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK.

Authorship of this paper should be cited as "International Wheat Genome Sequencing Consortium" (IWGSC, 2018).

\*Corresponding author. Email: rudi.appels@unimelb.edu.au (R.A.); eversole@eversoleassociates.com (K.E.); stein@ipk-gatersleben.de (N.S.)

†Major contributors

‡Working group leader(s) or co-leaders

## SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/361/6403/eaar7191/suppl/DC1](http://www.sciencemag.org/content/361/6403/eaar7191/suppl/DC1)

Materials and Methods

Figs. S1 to S59

Tables S1 to S43

References (56–186)

Databases S1 to S5

13 December 2017; accepted 11 July 2018

10.1126/science.aar7191

## RESEARCH ARTICLE SUMMARY

## WHEAT GENOME

# The transcriptional landscape of polyploid wheat

R. H. Ramírez-González\*, P. Borrill\*†, D. Lang, S. A. Harrington, J. Brinton, L. Venturini, M. Davey, J. Jacobs, F. van Ex, A. Pasha, Y. Khedikar, S. J. Robinson, A. T. Cory, T. Florio, L. Concia, C. Juery, H. Schoonbeek, B. Steuernagel, D. Xiang, C. J. Ridout, B. Chalhoub, K. F. X. Mayer, M. Benhamed, D. Latrasse, A. Bendahmane, International Wheat Genome Sequencing Consortium, B. B. H. Wulff, R. Appels, V. Tiwari, R. Datla, F. Choulet, C. J. Pozniak, N. J. Provart, A. G. Sharpe, E. Paux, M. Spannagl, A. Bräutigam, C. Uauy†

**INTRODUCTION:** Polyploidy, arising from whole-genome duplication or interspecific hybridization, is ubiquitous across the plant and fungal kingdoms. The presence of highly related genes in polyploids, referred to as homoeologs, has been proposed to confer adaptive plasticity—for example, through neofunctionalization of duplicated genes or tissue-specific expression. This plasticity has facilitated the domestication and adaptation of major polyploid crops (e.g., wheat, cotton, and coffee). However, despite its likely importance, we have a limited understanding of the effect of polyploidy on gene expression and the extent to which homoeologs are similar or different in their expression patterns across development and tissues.

**RATIONALE:** Bread wheat is a polyploid derived from the hybridizations between three distinct diploid species and is an informative system for analyzing the effects of recent polyploidy on gene expression. Understanding the coordination of homoeologs and identifying the mechanisms associated with these processes should help define strategies to improve trait biology in a crop that provides more than 20% of the protein and caloric intake of humans.

**RESULTS:** Here we leverage 850 wheat RNA-sequencing samples, alongside the annotated genome, to determine the similarities and differences between homoeolog expression across a range of tissues, developmental stages, and

cultivars. On average, ~30% of wheat homoeolog triads (composed of A, B, and D genome copies) showed nonbalanced expression patterns, with higher or lower expression from a single homoeolog with respect to the other two. These differences between homoeologs were associated with epigenetic changes affecting DNA methylation and histone modifications. Although nonbalanced homoeolog expression could be partially predicted by expression in diploid ancestors, large changes in relative homoeolog expression were observed owing to polyploidization.

Our results suggest that the transposable elements in promoters relate more closely to the variation in the relative expression of homoeologs across tissues than to a ubiquitous effect across all tissues. We found that homoeologs with the highest inter-tissue variation

had promoters with more frequent transposable element insertions and more varied cis-regulatory elements than homoeologs that were stable across tissues. We also identified

## ON OUR WEBSITE

Read the full article at <http://dx.doi.org/10.1126/science.aar6089>

expression asymmetry along wheat chromosomes. Homoeologs with the largest inter-tissue, inter-cultivar, and coding sequence variation were most often located in the highly recombinogenic distal ends of chromosomes. These transcriptionally dynamic homoeologs are under more relaxed selection pressure, potentially representing the first steps toward functional innovation through neo- or subfunctionalization.

We generated tissue- and stress-specific co-expression networks that reveal extensive coordination of homoeolog expression throughout development. These networks, alongside detailed gene expression atlases ([www.wheat-expression.com](http://www.wheat-expression.com) and <http://bar.utoronto.ca>), lay the groundwork to identify candidate genes influencing agronomic traits in wheat.

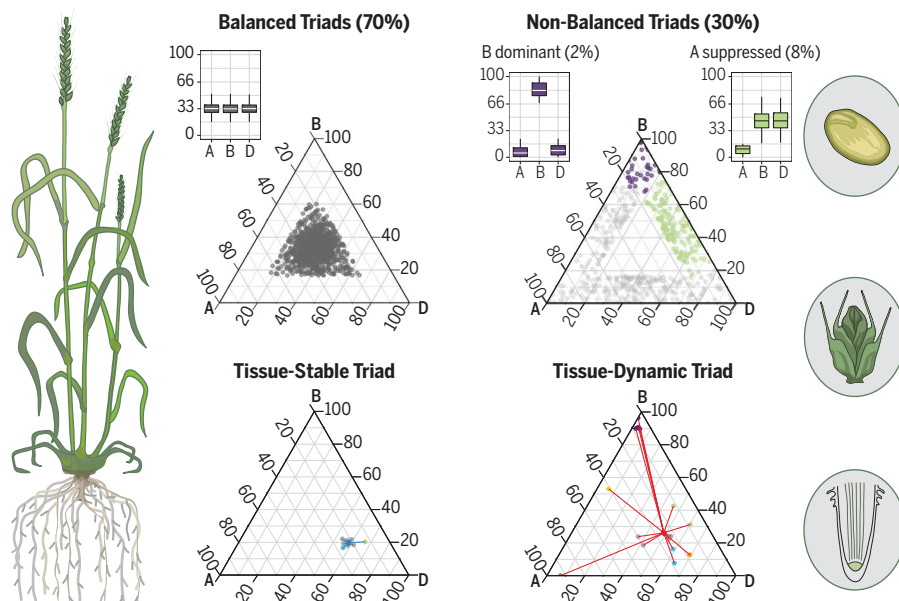
**CONCLUSION:** This study provides detailed insights into the transcriptional landscape of bread wheat, an evolutionarily young polyploid. Our work shows that homoeolog expression patterns in bread wheat have been shaped by polyploidy and are associated with both epigenetic modifications and variation in transposable elements within promoters of homoeologous genes. The extensive datasets and analyses presented here provide a framework that can help researchers and breeders develop strategies to improve crops by manipulating individual or multiple homoeologs to modulate trait responses. ■

The list of author affiliations is available in the full article online.

\*These authors contributed equally to this work.

†Corresponding author. Email: [cristobal.uauy@jic.ac.uk](mailto:cristobal.uauy@jic.ac.uk) (C.U.); [philippa.borrill@jic.ac.uk](mailto:philippa.borrill@jic.ac.uk) (P.B.)

Cite this article as R. H. Ramírez-González et al., *Science* **361**, eaar6089 (2018). DOI: [10.1126/science.aar6089](https://doi.org/10.1126/science.aar6089)



**Homoeolog expression patterns in polyploid wheat.** Seventy percent of triads (A, B, and D homoeologs) show balanced expression among homoeologs and are ubiquitously expressed (left), whereas ~30% show nonbalanced expression and are more tissue-specific (right; symbolized by three exemplar tissues). Variation in promoter elements and non-synonymous substitution rates distinguish between individual triads with stable relative expression across tissues and triads with more inter-tissue variation (tissue-dynamic triads).



## RESEARCH ARTICLE

## WHEAT GENOME

# The transcriptional landscape of polyploid wheat

R. H. Ramírez-González<sup>1\*</sup>, P. Borrill<sup>1\*†</sup>, D. Lang<sup>2</sup>, S. A. Harrington<sup>1</sup>, J. Brinton<sup>1</sup>, L. Venturini<sup>3</sup>, M. Davey<sup>4</sup>, J. Jacobs<sup>4</sup>, F. van Ex<sup>4</sup>, A. Pasha<sup>5</sup>, Y. Khedikar<sup>6</sup>, S. J. Robinson<sup>6</sup>, A. T. Cory<sup>7</sup>, T. Florio<sup>1</sup>, L. Concia<sup>8</sup>, C. Juery<sup>9</sup>, H. Schoonbeek<sup>1</sup>, B. Steuernagel<sup>1</sup>, D. Xiang<sup>10</sup>, C. J. Ridout<sup>1</sup>, B. Chalhoub<sup>11</sup>, K. F. X. Mayer<sup>2,12</sup>, M. Benhamed<sup>8</sup>, D. Latrasse<sup>8</sup>, A. Bendahmane<sup>8</sup>, International Wheat Genome Sequencing Consortium<sup>13†</sup>, B. B. H. Wulff<sup>1</sup>, R. Appels<sup>14</sup>, V. Tiwari<sup>15</sup>, R. Datla<sup>10</sup>, F. Choulet<sup>9</sup>, C. J. Pozniak<sup>7</sup>, N. J. Provart<sup>5</sup>, A. G. Sharpe<sup>16</sup>, E. Paux<sup>9</sup>, M. Spannagl<sup>2</sup>, A. Bräutigam<sup>17§</sup>, C. Uauy<sup>1†</sup>

The coordinated expression of highly related homoeologous genes in polyploid species underlies the phenotypes of many of the world's major crops. Here we combine extensive gene expression datasets to produce a comprehensive, genome-wide analysis of homoeolog expression patterns in hexaploid bread wheat. Bias in homoeolog expression varies between tissues, with ~30% of wheat homoeologs showing nonbalanced expression. We found expression asymmetries along wheat chromosomes, with homoeologs showing the largest inter-tissue, inter-cultivar, and coding sequence variation, most often located in high-recombination distal ends of chromosomes. These transcriptionally dynamic genes potentially represent the first steps toward neo- or subfunctionalization of wheat homoeologs. Coexpression networks reveal extensive coordination of homoeologs throughout development and, alongside a detailed expression atlas, provide a framework to target candidate genes underpinning agronomic traits in wheat.

**P**olyploidy arises from whole-genome duplication or interspecific hybridization and is ubiquitous in eukaryotic plant and fungal lineages. Polyploidy has been proposed to confer adaptive plasticity, thereby shaping the evolution of plants, fungi, and, to a lesser degree, animals (1, 2). This plasticity has facilitated the domestication and adaptation of several major crop species (3), including hexaploid bread wheat (*Triticum aestivum*; AABBDD subgenome), which is derived from relatively recent interspecific hybridizations between three different diploid species. In such polyploids, gene duplication alters the transcriptional landscape (4) by providing additional flexibility to adapt and evolve new patterns of gene expression for homoeologous gene copies (5). This flexibility has been suggested to be an important mechanism for controlling adaptive traits (6, 7)—for example, through neofunctionalization of dupli-

cated genes (8) or tissue-specific expression (9). However, despite the likely importance of polyploidy in affecting gene expression, we have a limited understanding of the extent to which homoeologs resemble or differ from each other in their expression patterns, the spatiotemporal dynamics of these relationships, and how epistatic interactions between individual homoeologs affect biological traits. The new genomic resources available for wheat (10), along with its meiotic stability (11) and syntenic gene order (12), make it a particularly informative system for gaining insight into the effects of recent polyploidy on gene expression.

In this study, we leveraged available RNA sequencing (RNA-seq) data (529 samples from 28 studies) and added 321 samples to explore global gene expression in hexaploid wheat across a diverse range of tissues, developmental stages, cultivars, and environmental conditions (13). We

organized these sets of RNA-seq samples into partially overlapping datasets from (i) a single developmental time course experiment ( $n = 209$  samples), (ii) the reference accession Chinese Spring (CS) under nonstress conditions ( $n = 123$  samples), (iii) four main tissue types under nonstress conditions ( $n = 537$  samples), and (iv) seedling samples from abiotic ( $n = 50$ ) and biotic ( $n = 163$ ) stress experiments including controls (table S1). These datasets, alongside a complete and annotated genome and transcriptome (10), provide an opportunity to conduct homoeolog-specific transcriptome profiling and to generate gene regulatory networks to better understand the spatiotemporal coordination of individual homoeologs underlying trait biology on a genome-wide scale.

## A developmental gene expression atlas in polyploid wheat

We first assessed expression patterns through a developmental time course of the commercial wheat cultivar Azhurnaya, including 209 RNA-seq samples representing 22 tissue types from grain, root, leaf, and spike samples across multiple time points (Fig. 1). We quantified expression using pseudoalignment of RNA-seq reads to the RefSeqv1.0 transcriptome, as implemented in kallisto (14), which accurately quantifies reads in a homoeolog-specific manner in polyploid wheat (13, 15) (figs. S1 and S2). We found evidence of expression for 83,741 (75.6% of 110,790) high-confidence genes, on the basis of expression of >0.5 transcripts per million (TPM) in at least one of the 22 tissue types, and we conducted complexity (table S2) and differential expression analyses (fig. S3). Tissue type distinguished samples across development (fig. S4) (13), consistent with observations in other plant and animal species (16, 17). Within similar tissue types, subgenome of origin also influenced expression patterns, consistent with previous results in wheat grain samples (18). This gene expression atlas provides a valuable resource for breeders and researchers to query for and analyze their genes of interest through [www.wheat-expression.com](http://www.wheat-expression.com) (15) and the Wheat eFP Browser at [http://bar.utoronto.ca/efp\\_wheat/cgi-bin/efpWeb.cgi](http://bar.utoronto.ca/efp_wheat/cgi-bin/efpWeb.cgi) (fig. S5) (19).

## Homoeolog expression patterns

In polyploid wheat, quantitative variation for many agronomic traits is modulated by genetic interactions between multiple sets of homoeologs in the A, B, and D subgenomes (20). These

<sup>1</sup>John Innes Centre, Norwich Research Park, NR4 7UH Norwich, UK. <sup>2</sup>Plant Genome and Systems Biology, Helmholtz Center Munich, Ingolstaedter Landstrasse 1, 85764 Neuherberg, Germany. <sup>3</sup>Earlham Institute, Norwich Research Park, NR4 7UZ Norwich, UK. <sup>4</sup>Bayer Crop Science, Innovation Center, Technologiepark 38, 9052 Zwijnaarde, Belgium. <sup>5</sup>Department of Cell and Systems Biology, Centre for the Analysis of Genome Evolution and Function, University of Toronto, 25 Wilcocks Street, Toronto, ON M5S 3B2, Canada. <sup>6</sup>Saskatoon Research and Development Centre, Agriculture and Agri-Food Canada, 107 Science Place, Saskatoon, SK S7N 0X2, Canada. <sup>7</sup>Crop Development Centre, University of Saskatchewan, Agriculture Building, 51 Campus Drive, Saskatoon, SK S7N 5A8, Canada. <sup>8</sup>Institut de Plant Sciences Paris-Saclay (IP2S), UMR 9213/UMR1403, CNRS, INRA, Université Paris-Sud, Université d'Evry, Université Paris-Diderot, Sorbonne Paris-Cité, Bâtiment 630, 91405 Orsay, France. <sup>9</sup>GDEC, INRA, UCA, 5 Chemin de Beaulieu, Clermont-Ferrand 63039, France. <sup>10</sup>Aquatic and Crop Resource Development, National Research Council Canada, 110 Gymnasium Place, Saskatoon, SK S7N 0W9, Canada. <sup>11</sup>INRA, 2 rue Gaston Crémieux, Evry 9057, France. <sup>12</sup>School of Life Sciences Weihenstephan, Technical University Munich, Munich, Germany. <sup>13</sup>IWGSC, 5207 Wyoming Road, Bethesda, MD 20816, USA. <sup>14</sup>School of BioSciences, University of Melbourne, AgriBio, La Trobe University, and School of Veterinary and Life Sciences, Murdoch University, 90 South Street, Perth, WA 6150, Australia. <sup>15</sup>Plant Sciences and Landscape Architecture, University of Maryland, 4291 Field House Drive, College Park, MD 20742, USA. <sup>16</sup>Global Institute for Food Security, University of Saskatchewan, 110 Gymnasium Place, Saskatoon, SK S7N 4J8, Canada. <sup>17</sup>Molecular Genetics, IPK Gatersleben, Corrensstrasse 3, 06466 Gatersleben, Germany.

\*These authors contributed equally to this work.

†Corresponding author. Email: [cristobal.uauy@jic.ac.uk](mailto:cristobal.uauy@jic.ac.uk) (C.U.); [philippa.borrill@jic.ac.uk](mailto:philippa.borrill@jic.ac.uk) (P.B.). ‡IWGSC collaborators and affiliations are listed in the supplementary materials. §Present address: Institute for Computational Biology, Faculty of Biology, University of Bielefeld, 33501 Bielefeld, Germany.

interactions range from buffering effects observed when gene homoeologs are functionally redundant (21) to dominance effects where variation in a single homoeolog can lead to dominant phenotypes (22). Understanding how these interactions influence gene expression will help inform strategies to improve crops by targeting and manipulating individual or multiple homoeologs to quantitatively modulate trait responses (20).

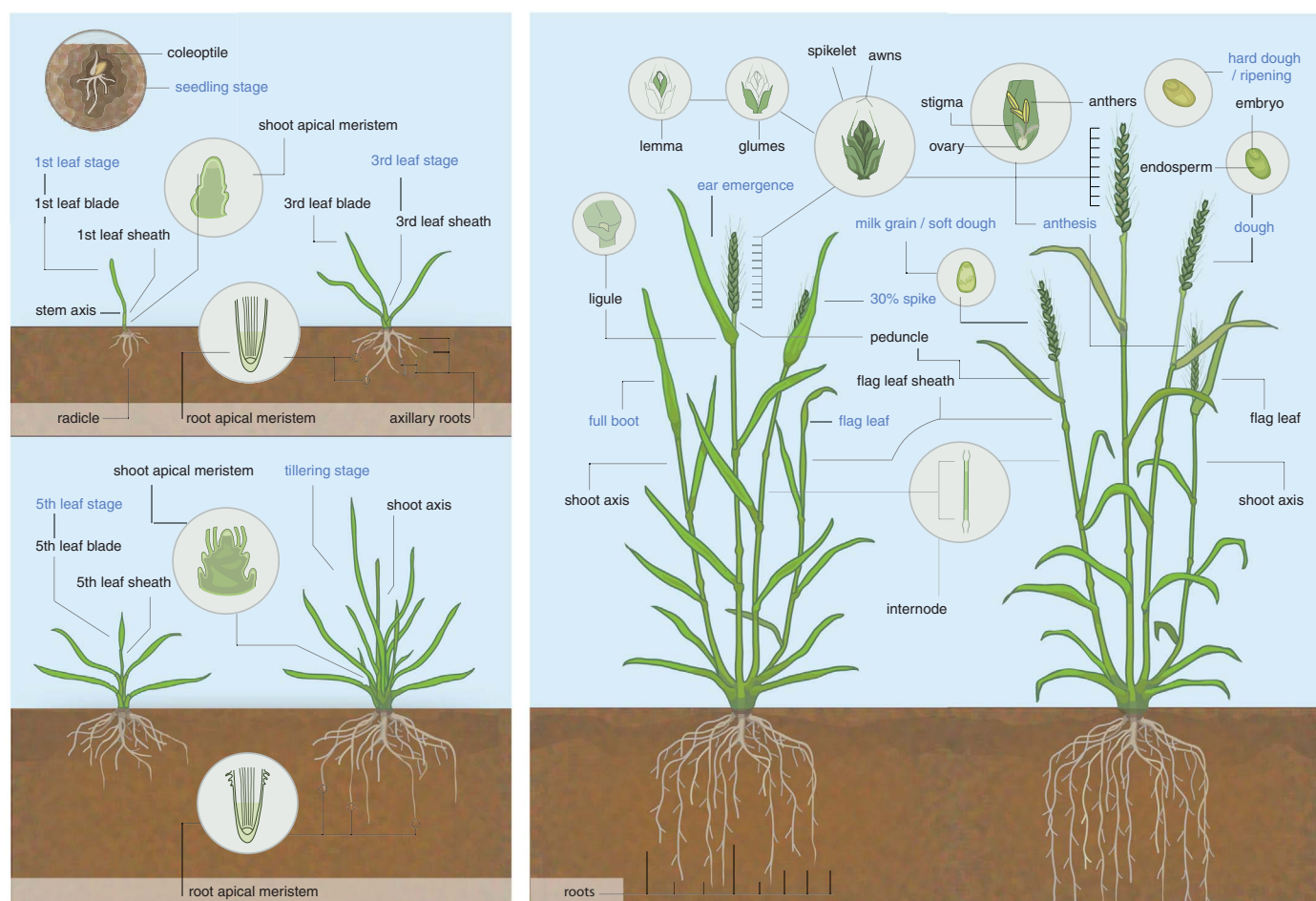
To determine patterns of homoeolog expression, we analyzed 123 RNA-seq samples representing 15 tissues under nonstress conditions (table S1) from CS. This was the same accession used to generate the reference genome (10); thus, cultivar-specific polymorphisms were excluded from our analysis. We found evidence of expression for 82,567 (74.5%) high-confidence genes, consistent with the developmental time course of the cultivar Azhurnaya. We focused on 53,259 genes that had a 1:1:1 correspondence across the three homoeologous subgenomes, referred to as triads, and a summed expression of >0.5 TPM across the triad (64.5% of expressed genes, 96.1% of all triads; table S3). The majority of these expressed triads (94.3%) were in ancestral (i.e., syntenic) physical positions in at

least two of the three subgenomes [50,238 genes corresponding to 16,746 syntenic triads (10)], whereas 5.7% (1007 triads) had all genes in non-syntenic positions and are thus referred to as nonsyntenic triads. For each of the 17,753 expressed triads, we standardized the relative expression of the A, B, and D subgenome homoeologs (fig. S6) so that the sum was 1.0 in each individual tissue. In this way, the relative abundance of homoeolog expression is comparable within triads, as well as across tissues, allowing the study of homoeolog expression bias (23).

We performed a global analysis combining data across all 15 tissues and focused on the 16,746 syntenic triads, but we discuss patterns in nonsyntenic triads where relevant. We found that the D subgenome had a subtly yet significantly higher relative abundance (33.65%) than the B (33.29%) and A (33.06%) subgenomes (Kruskal-Wallis  $P < 0.001$ ; tables S4 and S5). The homoeolog expression bias of the D subgenome is unlikely to reflect technical issues (fig. S2) and was found in 11 of the 15 tissues, was consistent across multiple expression abundance cutoffs, and was also significant in the developmental time course (in all 22 tissues) of the cultivar Azhurnaya [figs. S7 to S9 and table S5 (13)]. This effect, however,

is subtle when compared with the homoeolog expression bias observed in evolutionarily older polyploid crops such as cotton, in which genome doubling occurred at an earlier time point (24).

The relative expression of each homoeolog determined a triad's position in the ternary plot for the global analysis (Fig. 2A), as well as for analyses of individual tissues (figs. S6 and S10). From these plots, we defined seven homoeolog expression bias categories (13): a balanced category, with similar relative abundance of transcripts from the three homoeologs, and six homoeolog-dominant or homoeolog-suppressed categories, classified on the basis of the higher or lower abundance of transcripts from a single homoeolog with respect to those from the other two (Fig. 2A). Most syntenic triads (72.5%) were assigned to the balanced category within each tissue, with balanced triads ranging from 62.6% in the stigma and ovary to 78.9% in roots (Fig. 2B and table S6). Triads with single-homoeolog dominance were infrequent (7.1%; range among tissues, 4.7 to 11.3%), whereas syntenic triads classified as single-homoeolog-suppressed were more common (20.5%; range, 16.3 to 27.1%; Fig. 2B). These patterns shifted significantly in the 1007 nonsyntenic triads, which had fewer balanced



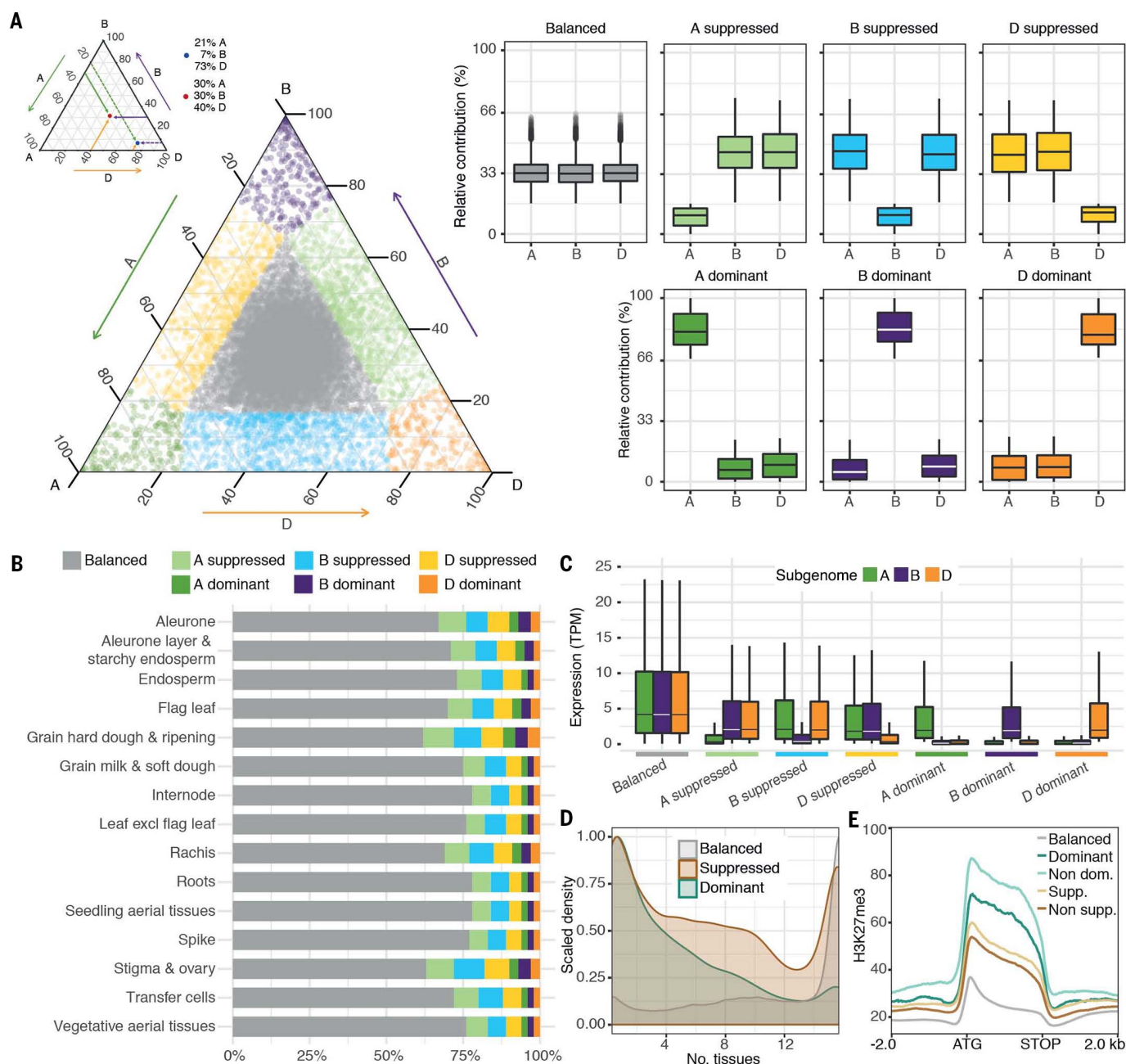
**Fig. 1. Developmental time course of bread wheat.** Shown is a schematic overview of tissues sampled for the RNA-seq expression atlas across multiple growth stages (labeled in blue). Details of all samples are provided in table S1.



triads (58.9%) and a higher proportion of dominant (14.5%) and suppressed (26.6%) triads across tissues ( $\chi^2 P < 0.001$ ; tables S7 and S8). Across tissues, no differences were observed in the frequency of single-homoeolog dominance between

subgenomes (tables S6 and S7). However, across all 15 tissues, D-homoeolog suppression was significantly less frequent (5.7%) than either A- or B-homoeolog suppression (7.5 and 7.2%, respectively; Kruskal-Wallis  $P < 0.05$ ), and this pat-

tern was also observed in nonsynthetic triads and the developmental time course (tables S6 to S8). This pattern explains in part the subtle homoeolog expression bias observed for the D subgenome relative to those observed for the



**Fig. 2. Homoeolog expression bias in syntenic homoeolog triads.**

(A) Ternary plot showing relative expression abundance of 16,746 syntenic triads (50,238 genes) in hexaploid wheat in the combined analysis of 15 tissues from Chinese Spring. Each circle represents a gene triad with an A, B, and D coordinate consisting of the relative contribution of each homoeolog to the overall triad expression (an example is shown on the top left). Triads in vertices correspond to single-subgenome-dominant categories, whereas triads close to edges and between vertices correspond to suppressed categories. Balanced triads are shown in gray. Box plots indicate the relative contribution of each subgenome based on triad assignment to the seven categories.

(B) Proportion of triads in each category of homoeolog expression bias across the 15 tissues (excl. excluding). (C) Box plot of absolute TPM expression abundance for each subgenome from the seven categories. (D) Number of tissues in which homoeolog-suppressed (brown), homoeolog-dominant (teal), and balanced (gray) triads are expressed. (E) Metagenome profile for histone H3K27me3 marks from -2 kb upstream of the ATG to +2 kb downstream of the stop codon (normalized for gene length) for balanced triads (gray), dominant triads separated into dominant (teal) and nondominant (pale blue) homoeologs, and suppressed triads separated into suppressed (tan) and nonsuppressed (brown) homoeologs.

A- and B-subgenome homoeologs. This observation is consistent with the lower distribution of repressive H3K27me3 (histone H3 lysine 27 trimethylation) histone marks across the gene body of D-subgenome homoeologs compared with those of the A- and B-subgenome homoeologs (fig. S11).

Genes from syntenic triads in the balanced category were expressed across a wider range of tissues and had higher absolute transcript abundance, on a per-subgenome basis (12.2 tissues; median, 4.03 TPM), than genes in the suppressed (9.1 tissues; median, 1.51 TPM) or dominant (6.9 tissues; median, 0.57 TPM) categories (two-sample Kolmogorov–Smirnov test,  $P < 0.001$ ; Fig. 2, C and D, and tables S9 and S10). The absolute transcript abundance data show that dominant triads are not the result of an overall increase in expression of a single homoeolog, but rather result from the relatively lower expression of the two other homoeologs.

To determine if the differences among homoeologs are a consequence of polyploidization, we analyzed RNA-seq data from diploid and tetraploid progenitor species and newly created synthetic hexaploid wheat (SHW) lines (25). We found that 67.5% of nonbalanced triads in modern-day wheat have a different homoeolog expression bias category than that observed in SHW, with all three subgenomes being equally affected (table S11 and figs. S12 to S14). Likewise, 47.1% of nonbalanced triads in SHW are in a different category than would be expected on the basis of the progenitor species, with D-subgenome homoeologs most strongly influenced (13) (table S12 and fig. S15). These results suggest that the polyploid context and the polyploidization process itself affect the relative expression of homoeologs compared with the baseline expression in the progenitor species (13), which has also been observed during the evolution of polyploid cotton (26) and monkeyflower (27).

We hypothesized that epigenetic mechanisms might be associated with differences in homoeolog expression patterns. To test this, we examined the associations of transposable elements (TEs), DNA methylation, and histone modifications with the relative expression of triads in leaves of CS. We found no clear relationship between the presence of TEs in promoter regions and altered expression patterns between homoeologs in dominant and suppressed triads (Tukey's Honestly Significant Difference  $P > 0.6$ ; fig. S16 and table S13) (13). However, we identified significant differences in gene-body DNA methylation and histone modifications among homoeologs (13).

Gene-body CG methylation is widely conserved in angiosperms, although its functional significance is currently under debate (28, 29), given that two angiosperm species lack this epigenetic mark altogether (30). We found higher gene-body CG methylation in constitutively expressed triads than in more tissue-specific triads (balanced > suppressed > dominant; fig. S17). Within the nonbalanced triads, homoeologs with higher expression had higher CG methylation than their corresponding nondominant and sup-

pressed homoeologs (Mann-Whitney  $P < 0.001$ ; fig. S17). These results are consistent with gene-body CG methylation associated with housekeeping genes and its suggested role in homeostatic gene expression (29). Similarly, the more highly expressed homoeologs within nonbalanced triads had higher active (H3K36me3 and H3K9ac; ac, acetylation) histone marks and lower repressive (H3K27me3) histone marks in the gene body (Mann-Whitney  $P < 0.001$ ; fig. S11). For H3K27me3, these differences were not limited strictly to the gene body but extended into the upstream and downstream regions for both dominant and suppressed triads (Fig. 2E), consistent with the tight association of H3K27me3 with inactive gene promoters (31). These results suggest that epigenetic status in gene bodies, as well as upstream and downstream regions, is associated with homoeolog expression bias in polyploid wheat, consistent with results in monkeyflower showing changes in DNA methylation upon polyploidization (27).

Breeders rely on recombination to generate new combinations of haplotypes for improving cultivars. In wheat, chromosome position strongly influences recombination rates, with relatively low recombination rates in the interstitial and proximal regions (R2a, C, and R2b genomic compartments) but markedly higher rates toward the distal ends of the chromosomes (R1 and R3 genomic compartments) (32). In our analyses, syntenic triads in the balanced category were overrepresented in the low-recombination regions (R2 and C), which have higher levels of active histone marks (H3K36me3 and H3K9ac) (10), consistent with the higher expression of balanced triads. Homoeolog-dominant and homoeolog-suppressed triads were overrepresented toward the high-recombination distal ends of chromosomes (R1 and R3;  $\chi^2 P < 0.001$ ; table S14), which have higher levels of repressive (H3K27me3) histone marks (10), consistent with the lower expression of dominant and suppressed triads. This pattern was also observed in the developmental time course of the cultivar Azhurnaya. However, when comparing the CS and Azhurnaya cultivars (nine tissues in common), we found that 84.5% of genes in the R2 and C regions had the same expression category between cultivars, whereas only 72.2% of genes in the R1 and R3 regions did so ( $\chi^2 P < 0.001$ ; table S15). These differences in homoeolog expression bias across cultivars have important implications for breeding because they suggest that through genetic crosses, breeders not only generate new combinations of haplotypes with differential expression of alleles, but also rearrange and select for homoeolog expression bias between cultivars.

### Variation of triad expression patterns

Polyploidy may confer phenotypic plasticity by allowing homoeologs to be expressed differently across tissues and/or environmental conditions (8). Our analyses above provide a static overview of the relative homoeolog expression bias in individual tissues. Therefore, we explored whether syntenic triads retain their homoeolog expres-

sion bias category across the 15 tissues (table S16). We found that 83.6% of balanced triads remained balanced in each of the 15 individual tissues, whereas dominant and suppressed triads tended to be more variable across tissues, with only 73.4 and 62.2%, respectively, staying within their global dominance group across all 15 tissues (Fig. 3A). Dominant and suppressed triads shifted most often to adjacent categories (16 to 20%) in the ternary plots and in few cases (<3.0%) changed to opposite categories (fig. S18). These patterns were also observed in the developmental time course (table S16). These data show that across tissues, triads most often remained consistent in their homoeolog expression bias classification, a phenomenon also seen across seven tissues in allotetraploid *Tragopogon mirus* (33).

To complement this analysis, we determined the variation in behavior of each triad within the ternary plot across the 15 tissues by calculating the mean distance between the triad's position in each tissue and its global average position (13) (fig. S19). This generated a distribution of mean distances (Fig. 3B); we focused on the 10% most stable triads (defined as those having the shortest mean distances across tissues) and the 10% most dynamic triads (largest mean distances) (Fig. 3, B to D). Stable triads were expressed more highly than dynamic triads (median, 8.2 versus 3.2 TPM;  $P < 0.001$ ) and had a higher expression breadth, being expressed across almost all samples, whereas dynamic triads were more tissue-specific ( $P < 0.001$ ) (Fig. 3E and table S17). Stable triads were enriched for high-level gene ontology (GO-slim) terms associated with housekeeping processes (e.g. translation and cell cycle), whereas dynamic triads were enriched for defense and external stimuli responses and secondary metabolic processes, functions that more frequently determine differences in individual fitness (table S18) (6). In the global analysis, stable triads were significantly enriched for the balanced category (94.2%), whereas dynamic triads were almost equally spread between suppressed (37.9%), dominant (30.5%), and balanced (31.6%) categories ( $\chi^2 P < 0.001$ ; Fig. 3F and table S19). This pattern is consistent with stable triads being most frequently located in proximal regions (C), whereas dynamic triads tend to locate in distal ends of chromosomes (R1 and R3) ( $\chi^2 P < 0.001$ ; table S20). These results demonstrate expression asymmetry across wheat chromosomes, whereby the high-recombination distal ends of chromosomes have triads that exhibit higher homoeolog expression bias, are more dynamic across tissues, and have higher expression variation between cultivars than triads in the low-recombination proximal regions (Fig. 3G). This asymmetry is also reflected in the contrasting distributions of histone marks and DNA methylation along chromosomes (10). The difference in epigenetic marks identified in leaves makes it tempting to speculate that epigenetic marks may also be associated with triad expression variation across multiple tissues.

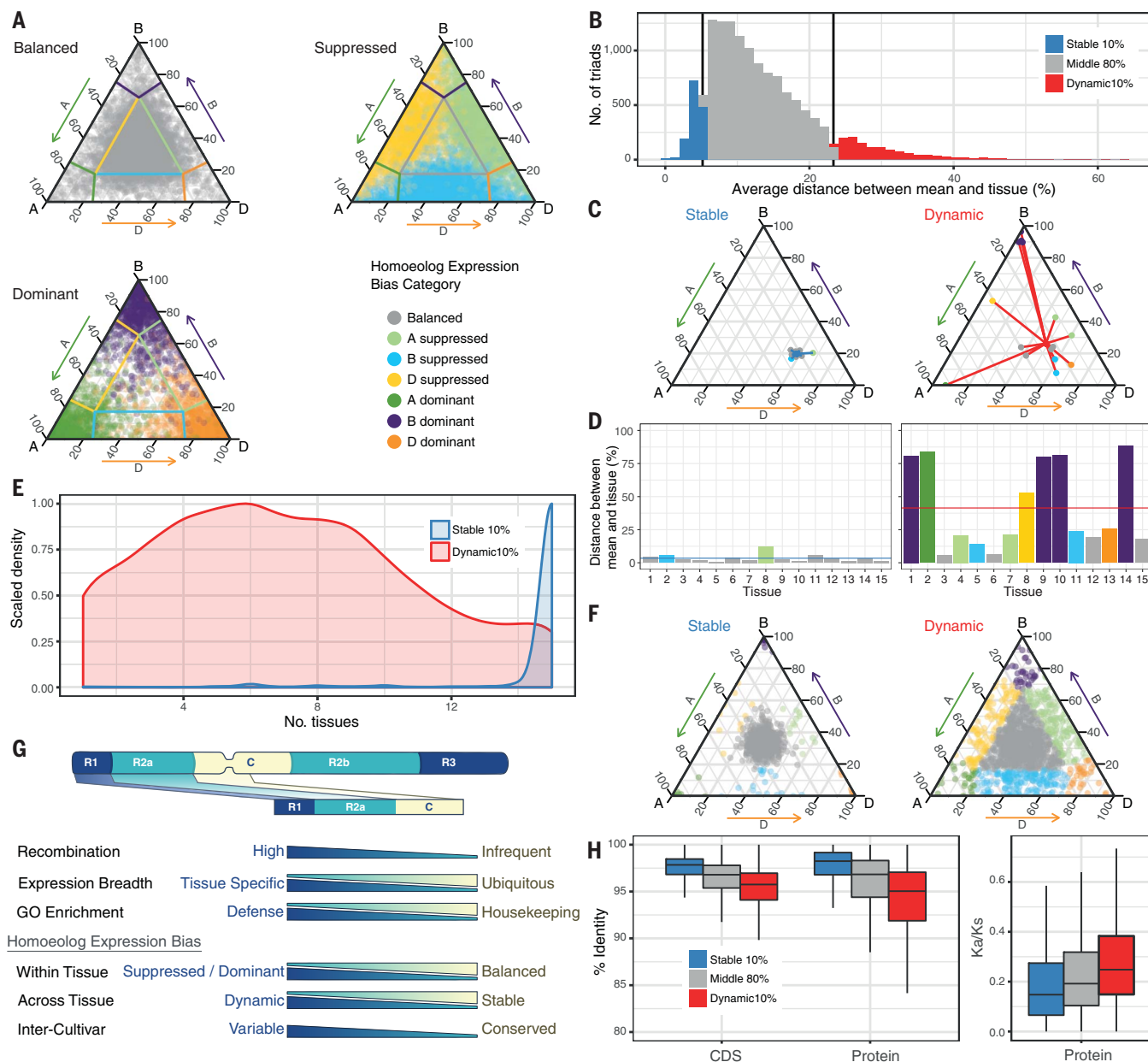
We next investigated if divergence in spatial expression patterns (as measured by the mean



distance statistic described above) was coupled with 5' promoter and protein sequence divergence in syntenic triads (13). The 1.5-kilobase (kb) promoters of dynamic triads more frequently contained TEs (88.3 versus 79.2%), which were closer to the translation start site (1113 versus 1234 bp away) but shorter (median, 220 versus 259 bp), than those in stable triads, leading to equivalent TE densities (all comparisons, Kruskal-Wallis  $P < 0.001$ ; fig. S20 and table S13). These

closer, more frequent, and shorter TEs could potentially act as novel cis-regulatory elements (34) or influence epigenetic marks (35). These results indicate that the promoter TE landscape relates more closely to the variation in the relative expression of homoeologs across tissues than to a ubiquitous effect across all tissues (table S13). Although only subtle differences in sequence identity were found between stable and dynamic triads (85.5 versus 85.0%;  $P = 0.045$ ) (Fig. 3H

and table S21), dynamic triads had fewer conserved transcription factor (TF) binding site motifs across the three homoeologs (37% fewer;  $P < 0.001$ ; fig. S21). Across coding sequences, we showed a stepwise decrease in conservation of both the nucleotide and protein identities from stable (average, 97.2% coding sequence and 97.3% protein) to dynamic (95.0 and 93.4%) triads (both  $P < 0.001$ ; table S21). We compared nonsynonymous ( $K_a$ ) with synonymous ( $K_s$ ) substitution



**Fig. 3. Variation of triad expression patterns.** (A) Variation of balanced, dominant, and suppressed triads (assigned on the basis of global analysis) across 15 tissues. (B) Distribution of mean distance of triad variation across 15 tissues for 14,258 triads expressed in at least six tissues. The 10% most stable (blue) and 10% most dynamic (red) triads were defined. (C) Ternary plots of representative stable and dynamic triads and (D) bar graphs of the distance between the triad position in the 15 individual tissues and the triad global average position (horizontal line).

Color-coding is as in (A). (E) Number of tissues in which stable (blue) and dynamic (red) genes are expressed (table S36). (F) Homoeolog expression bias classification of stable and dynamic triads in global analysis. (G) Schematic representation of a wheat chromosome based on genomic compartments and features associated with distal (R1 and R3) and interstitial and proximal (R2 and C) regions. (H) Box plots of percent coding and protein sequence identity (left) and  $K_a/K_s$  ratio (right) for stable 10% (blue), middle 80% (gray), and dynamic 10% (red) triads.

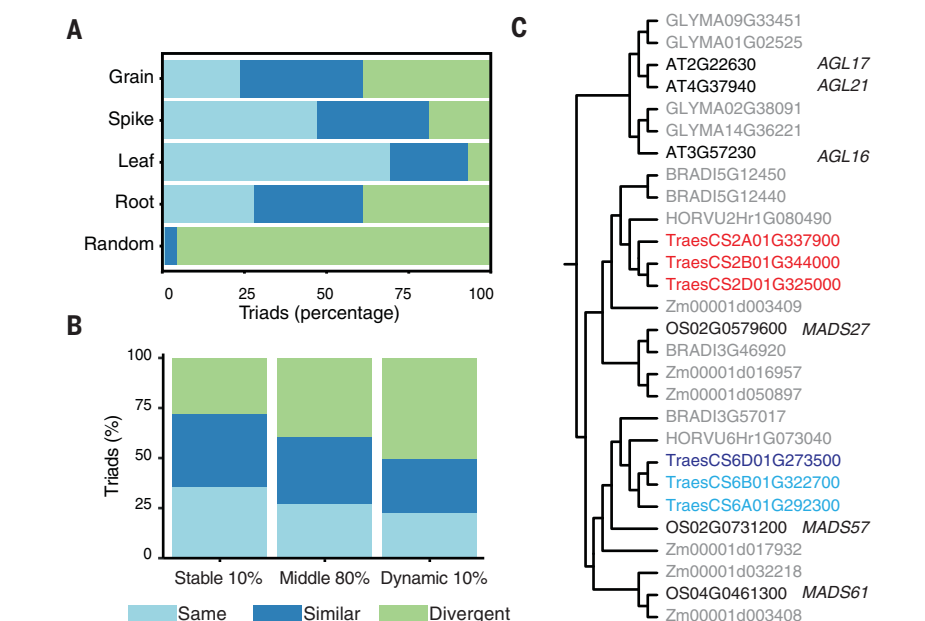
rates between homoeologs and observed that dynamic triads had significantly higher  $K_a/K_s$  than stable triads (0.33 versus 0.21; Mann-Whitney  $P < 0.001$ ; Fig. 3H and table S22). This higher ratio suggests that triads with greater divergence in spatial expression patterns are under more relaxed selection pressure, as seen for duplicated genes in humans (9), but not in soybean (36) and carp (37). This conclusion is supported by the observation that nonsynthetic triads, which had greater expression divergence (10.5% larger mean distance; Mann-Whitney  $P < 0.001$ ), also had significantly higher  $K_a/K_s$  (0.42; Mann-Whitney  $P < 0.001$ ) compared with syntenic triads (table S22). The above relationships were consistent when using different percentage cutoffs to define stable and dynamic triads (5 and 25%), as well as in the developmental time course of the cultivar Azhurnaya (tables S21 and S22). These results show positive coupling of divergence in spatial expression patterns with divergence in TE and cis-regulatory elements in promoters and sequence divergence in coding sequence among wheat homoeologs. It is possible that divergence in spatial expression patterns, alongside relaxation of selection pressure, can lead to functional innovation through homoeolog neo- or subfunctionalization.

### Coordinated expression of homoeolog triads

Our analyses provide a framework to describe the relative expression of individual homoeologs between discrete triads in space and time. To understand how this coordination of homoeolog spatiotemporal expression may influence biological processes, we developed a series of coexpression networks to provide insight into tissue-specific developmental and stress-related processes.

We constructed four separate tissue-specific coexpression networks from nonstress RNA-seq samples from grain ( $n = 119$  samples), leaf ( $n = 245$ ), root ( $n = 45$ ), and spike ( $n = 128$ ), using all genes expressed at more than 0.5 TPM in the given tissue (13). These networks were composed of 51 to 78 modules and contained 42.3 to 88.0% of all expressed genes in each tissue (fig. S22, table S23, and data S1). We found that across all tissue networks, homoeologs from 37.4% of the syntenic triads were in the same coexpression module, suggesting a highly coordinated expression pattern for these triads ( $\chi^2 P < 0.0001$  with respect to random triads; table S24). However, the majority of triads (62.6%) had at least one homoeolog outside the same module.

To quantify whether homoeologs outside the module had similar or divergent expression patterns, we calculated a threshold based on the pairwise distance between homoeologs (13). We found that 29.6% of syntenic triads had a divergent pattern, wherein the expression of at least one homoeolog exceeded the distance threshold in the tissue network (Fig. 4A and fig. S23). Conversely, 33% of triads had a similar pattern, wherein all pairwise distances between homoeologs were lower than the threshold, suggesting a



**Fig. 4. Homoeolog coexpression patterns in tissue networks.** (A) Triad assignment to same, similar, and divergent modules in tissue coexpression networks. (B) Stable, middle, and dynamic triad assignment to same, similar, and divergent modules in the root network. (C) Neighbor-joining phylogenetic tree of homologs for the *Arabidopsis* MADS\_II gene AGL21. Wheat orthologs from chromosome group 2 are assigned to the root-specific module 61 (red), whereas chromosome group 6 orthologs are assigned to modules 1 and 13 (blue and purple).

subtler variation in a single homoeolog. These values showed significant variation between tissue networks, ranging from 7% divergent triads in the leaf network to more than 38% divergent triads in the root and grain networks (Fig. 4A). Nonsynthetic triads had a higher proportion of divergent triads in all tissue networks compared with syntenic triads (mean, 35.1 versus 29.6%;  $\chi^2 P < 0.001$ ; table S24). Using the same criterion as before (triad mean distance between tissues and global average position), we identified the 10% most stable and dynamic syntenic triads for each tissue-specific network. We found that dynamic triads were more frequently in divergent modules than stable triads for all four tissue networks ( $P < 0.001$ ; Fig. 4B and fig. S24). These results are consistent with the homoeolog expression bias analyses and support the idea that although many triads are expressed in a coordinated spatiotemporal pattern (with the same or similar profile), almost 30% of syntenic and 35% of nonsynthetic triads have a divergent expression profile. Transcriptional divergence occurs both immediately upon polyploidization and after polyploidization (figs. S12 to S14) and may represent initial steps toward neo- or subfunctionalization of wheat homoeologs.

### Exploiting development and stress networks for biological discovery

To explore the potential for biological discovery, we first compared modules between networks to identify tissue-specific gene networks. Across the

four networks, 73.2% of modules had significant overlap (Fisher's exact test,  $P < 0.05$ ) with modules in all four networks, with the root having the fewest conserved modules (61.1%) and the spike having the most (86.2%) (data S2). In the root, there were three modules that were not found in any other tissue, with the largest of these (root module 61; 82 genes) enriched for root-related plant ontology (PO) terms (e.g., root procambium,  $P = 3.3 \times 10^{-5}$ , and central root cap of primary root,  $P = 4.5 \times 10^{-5}$ ; table S25). We hypothesized that genes encoding TFs controlling processes related to these PO terms would also be coexpressed within module 61. We found that four of the 10 genes encoding TFs in this root-specific module had known functions related to root development in *Arabidopsis* or rice (38, 39) (table S26). Three of these TFs belonged to one homoeolog triad in the MADS\_II family, and one of their *Arabidopsis* orthologs (AGL21) has been shown to regulate lateral root development through auxin accumulation (39). To understand the target genes of these TFs in wheat, we conducted a complementary network analysis using genie3, which predicted target genes of TFs across all 850 samples (13). Target genes of the three TFs were enriched for cell wall processes and lignification, consistent with their putative role in the differentiation zone where lateral roots emerge (tables S27 and S28). Closely related paralogs on chromosome group 6 in wheat were not located in root module 61; rather, they were in modules 1 and 13 (Fig. 4C). These modules were conserved in all other tissue networks, implying a more



general function for genes within them. Supporting this hypothesis, the rice ortholog of the chromosome 6 paralogs (*OsMADS57*) has been shown to play a role in tillering (40).

A key challenge for wheat breeding is the selection of cultivars with tolerance to multiple stresses. Therefore, we focused on stress responses in seedlings and young vegetative plants, for which 10 independent studies with 12 distinct abiotic and disease stresses were available (table S1). We constructed gene coexpression networks for abiotic and disease stresses separately, including control samples from the same studies to allow for links between disease status and gene expression (13). We integrated the two networks to identify modules that might be common to both abiotic and disease responses. We found 84 pairs of modules between the two networks that had significantly overlapping gene content and were significantly correlated with both an abiotic and a disease stress (tables S29 to S31). The most significant overlap was between disease module 12 and abiotic module 2 ( $P = 1.3 \times 10^{-94}$ ), which shared 355 genes (Fig. 5A). These two modules had similar enrichment for GO-slim terms relating to signal transduction and response to stimulus (table S32), suggesting that they might perform similar biological functions.

Among the 355 shared genes, there were 16 encoding TFs, six of which have orthologs in rice or *Arabidopsis* with proven roles in abiotic or disease stress, and a further three have orthologs differentially expressed during stress in these species (table S33). Furthermore, on the basis of the genie3 analysis, 11 of the 16 TFs have targets that are enriched in stress responses, and seven have targets that are enriched simultaneously in biotic and abiotic stress responses (Fig. 5B). Of the genes encoding these TFs, two homoeo-

logs stood out as potential common regulatory components of abiotic and disease response: *TraesCS5A01G237900* and *TraesCS5B01G236400*, which encode heat shock factor (HSF) TFs. These two HSF TF-encoding genes were in the top 10 most central genes within disease module 12 (table S34), as measured by intramodule connectivity, a value strongly correlated with the influence of a gene on a phenotype (41). The 387 predicted targets of the TFs encoded by these two genes were frequently allocated to module 12 of the disease network (39.5%) and module 2 of the abiotic stress network (28.0%) (table S35). The *Arabidopsis* ortholog of these genes, *TBFI*, was originally identified for its role in pathogen defense response (42) and has been shown to play a key role in the transition from growth to defense (43), while also positively regulating acquired thermotolerance (44). Recently, a “*TBFI* cassette” including the promoter and 5′ leader region of *TBFI* was used to engineer broad-spectrum disease resistance in both *Arabidopsis* and rice without a fitness cost (45). The fact that *Arabidopsis TBFI* is functional in rice suggests that this regulatory mechanism may be conserved across species, making the wheat orthologs identified here promising targets for further studies. These and other highly connected genes (table S34) are strong candidates for controlling stress responses, and the functions of their orthologs support this hypothesis. These results demonstrate the power of the datasets and show that integrating gene networks from wheat, alongside phylogenetic relationships and knowledge of biological function in model species, can help identify candidate genes for further study in wheat.

### Concluding statement

This study provides detailed insight into the spatiotemporal transcriptional landscape of

polyploid wheat. We find evidence that the differences in relative expression among homoeologs observed in modern-day wheat may have been established both upon the polyploidization of wheat itself and during the subsequent 10,000 years of polyploidy; these differences may have been determined through epigenetic changes affecting both DNA methylation and histone modifications. We identified asymmetries along wheat chromosomes for a series of features relating to homoeolog expression bias with important implications for breeding. Our work provides a framework for the generation of hypotheses about biological function in polyploid wheat, which can now be experimentally tested using recent developments in sequenced mutant populations (46) and genome editing approaches (47). Ultimately, this knowledge will help researchers and breeders modulate allelic variation across homoeologs to improve quantitative traits in polyploid wheat. This is an urgent task for achieving global food security, given that wheat provides more than 20% of the protein and caloric intake (48) of humans.

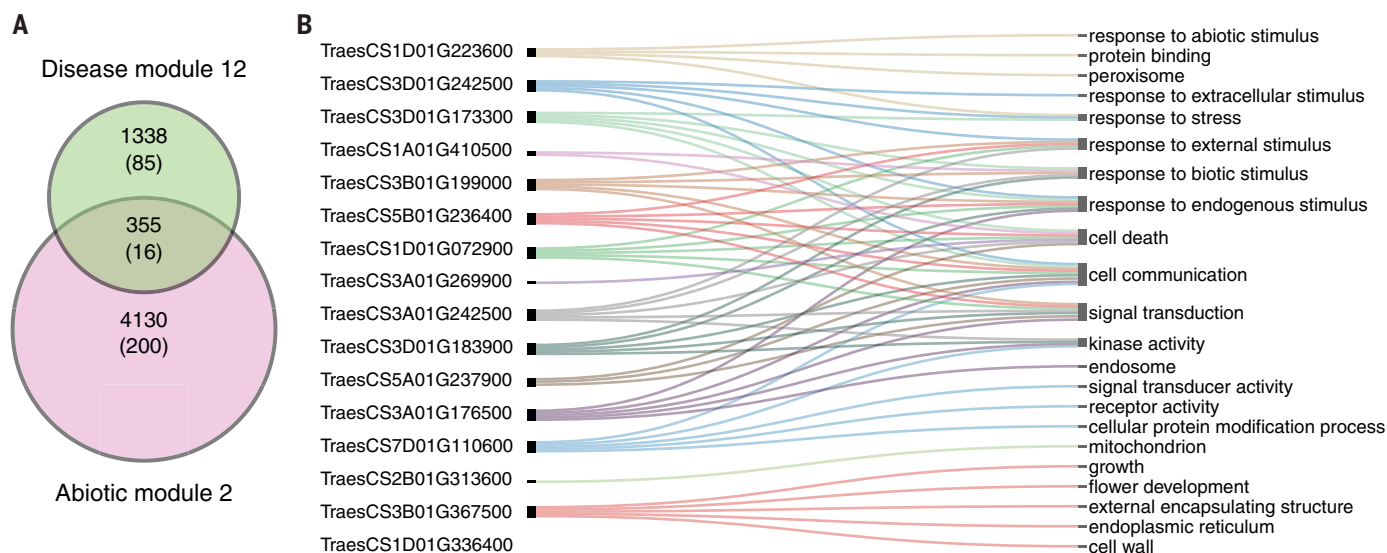
### Materials and methods

#### RNA-seq samples

We included 246 samples previously described (15) and complemented this with 283 RNA-seq samples which were deposited in the Short Reads Archive (SRA) between August 2015 and January 2017. An additional 321 RNA-seq samples from six studies were used for this analysis and are detailed in the supplementary materials (13).

#### Mapping of RNA-seq reads to reference

For all 850 samples, metadata was assigned as described in (15), with high and low-level factors for tissue, age, variety, and stress. Due to the relatively large number of low-level tissues (59), we



**Fig. 5. Overlapping modules within abiotic and disease stress networks. (A)** Number of genes in abiotic module 2 and disease module 12 and the overlap between modules. The number of transcription factors is indicated in parentheses. **(B)** Transcription factors found in both abiotic 2 and disease 12 (left) and the top five enriched GO terms of their targets, as identified by genie3 (right).

further defined an intermediate level of tissues comprising 32 factors (average 26.5; median 12 replicates per factor) which was used for this study. We also assigned an intermediate level of stress comprising 15 factors (average 14.5; median 6 replicates per factor). We used kallisto v0.42.3 (14) to map the 850 RNA-seq samples to the Chinese Spring RefSeqv1.0+UTR transcriptome reference. We used default parameters previously shown to result in accurate homoeolog-specific read mapping in polyploid wheat (15) (fig. S1). We summarized expression levels from the transcript level to the gene level using tximport v1.2.0. We established the criteria that at least 1% of samples for a given gene all required to have expression values over 0.5 TPM for that gene to be considered expressed (*initial 850 filter*).

To confirm that kallisto enables homoeolog specific mapping (15) we analyzed expression of HC genes expressed >0.5 TPM in nulli-tetrasomic wheat lines from the publicly available study SRP028357 (49). The nulli-tetrasomic lines were missing an entire chromosome (1A, 1B, or 1D) which was replaced by a duplication of another homoeologous chromosome, e.g. Nulli1ATetra1B has 0 copies of 1A, 2 copies of 1B and 1 copy of 1D. We determined stringent homoeolog-specific mapping using a series of criteria detailed in the supplementary materials (13).

### Analyses of expressed genes

Starting from the subset of genes considered expressed using the *initial 850 filter* criterion, we determined genes which were expressed in at least one tissue within the Azhurnaya developmental time course (209 samples; 22 intermediate tissues) and Chinese Spring no stress (123 samples; 15 intermediate tissues) datasets. For this analysis, we first calculated the average TPM expression of each gene in each of the intermediate tissue types (*average expression per tissue*). The number of samples that went into generating this *average expression per tissue* value varied for each intermediate tissue and are available in table S1. We considered a gene expressed when its *average expression per tissue* was > 0.5 TPM in at least one intermediate tissue. For both datasets we focused on HC gene models (10). Whilst expression data was also assessed for LC genes, we excluded these from the main analysis to avoid confounding effects from pseudogenes and low-quality gene models. Through this analysis we found evidence of expression for 83,741 (75.6%) HC genes in Azhurnaya and 82,567 (74.5%) HC genes in Chinese Spring.

Using the *average expression per tissue* values, we also determined the global expression of each gene across all tissues in which it was expressed (based on the >0.5 TPM criteria in the tissue). This generated an average value across tissues, rather than a geometric mean across all samples, to account for the variation in the number of samples per tissue. It also excludes tissues in which a gene is not expressed. This average across expressed tissues is referred to as either the “global analysis” or the “combined analysis

(all tissues)” across the main text and in the supplementary materials and tables.

### Relative expression levels of the A, B, and D subgenome homoeologs across triads

The analysis focused exclusively on the gene triads which had a 1:1:1 correspondence across the three homoeologous subgenomes, including 17,400 syntenic and 1074 nonsyntenic triads (total of 18,474 triads or 55,422 genes). Starting from the subset of genes considered expressed using the *initial 850 filter* criterion, we defined a *triad* as expressed when the sum of the A, B, and D subgenome homoeologs was > 0.5 TPM. This allowed us to include triads in which, for example, only a single homoeolog was expressed, and which could later be classified as a dominant triad. Using this criterion, we defined a total of 53,259 genes (17,753 triads) which were considered expressed (table S3). To standardize the relative expression of each homoeolog across the triad, we normalized the absolute TPM for each gene within the triad as follows

$$\text{expression}_A = \frac{\text{TPM}(A)}{\text{TPM}(A) + \text{TPM}(B) + \text{TPM}(D)}$$

$$\text{expression}_B = \frac{\text{TPM}(B)}{\text{TPM}(A) + \text{TPM}(B) + \text{TPM}(D)}$$

$$\text{expression}_D = \frac{\text{TPM}(D)}{\text{TPM}(A) + \text{TPM}(B) + \text{TPM}(D)}$$

where A, B, and D represent the gene corresponding to the A, B, and D homoeologs in the triad. The normalized expression was calculated for each one of the intermediate tissues and for the average across all expressed tissues (“*combined analysis*” as described previously). Fig. S6 shows an example of these calculations for the roots and the combined analysis across three triads. The values of the relative contributions of each subgenome per triad were used to plot the ternary diagrams using the R package ggtern (50).

### Definition of homoeolog expression bias categories

The ideal normalized expression bias for the seven categories was defined as shown in table S37.

We calculated the Euclidean distance (rdist function from R 3.3.2) from the observed normalized expression of each triad to each of the seven ideal categories listed above. We assigned the homoeolog expression bias category for each triad by selecting the shortest distance. This was done for each of the intermediate tissue as well as for the average across all expressed tissues (*combined analysis*).

### Analysis of the effects of polyploidy on homoeolog expression bias

We used RNA-seq data (25) which consisted of two datasets based on RNA-seq samples from the youngest leaf at fifth leaf stage. Dataset 1

(SHW1) included samples from tetraploid (BBAA) *Triticum turgidum* ssp. *turgidum* wheat accession AS2255, diploid *Ae. tauschii* (DD) accession AS60, and the synthetic hexaploid wheat (SHW1; BBAAADD) resulting from the cross between the tetraploid and *Ae. tauschii* accessions. Dataset 2 (SHW2) consisted of tetraploid *T. turgidum* ssp. *durum* cv Langdon (BBAA), the same diploid *Ae. tauschii* (DD) accession AS60, and an independent synthetic hexaploid wheat (SHW2) derived from Langdon x AS60 (BBAAADD). Note that AS2255 and Langdon are both *T. turgidum* ssp., but are defined as different subspecies based primarily on morphological features. These experiments recreate the polyploidization events that gave rise to modern bread wheat and the resulting SHW has the same genome composition as the CS and Azhurnaya datasets examined in this study.

We analyzed the RNA-seq from both datasets by mapping reads to the CS RefSeqv1.0 transcriptome using the same bioinformatics pipeline as before (see “Mapping of RNA-seq reads to reference” section). However, for the tetraploid datasets we used only the A and B subgenome transcripts as a reference, for the diploid D genome datasets we used only D subgenome transcripts, and for the SHW datasets we used the complete RefSeqv1.0 transcriptome as the reference, as in CS and Azhurnaya. To generate the *expected* hexaploid wheat transcriptome based on progenitor species we weighted the TPM values from the tetraploid by 2/3 and the AS60 TPM values by 1/3 to maintain a total TPM of  $10^6$  in the combined dataset. The *in-silico* hexaploid wheat generated from the weighted tetraploid and diploid TPM values (referred to hereafter as the “*expected*” *in-silico* dataset) allows the direct comparison with the *observed* TPM values in SHW. We defined the seven homoeolog expression bias categories for both the *expected in-silico* and the *observed* SHW transcriptomes using the same methods as for CS and Azhurnaya and compared the classification of triads between the *observed* and *expected* datasets (table S12). We next compared classifications to modern-day bread wheat CS and Azhurnaya. To enable a meaningful comparison across similar tissues from the Hao et al study (25) we used nine samples from the *PAMP Triggered Immune Response* dataset from CS and six samples from the Azhurnaya dataset (table S1). As before, we defined the seven homoeolog expression categories for the defined CS and Azhurnaya datasets and compared them with the SHW and the *in-silico* classifications (table S11).

### DNA methylation plant material and library preparation

Plants were grown as described in the *Chinese Spring tissues* study. The frozen leaves from the five samples at 3-leaf stage (Zadok stage 13) were ground and divided as input for the preparation of both RNA-seq libraries (detailed in *Chinese Spring tissues* study) and whole genome bisulfite sequencing (WGBS) libraries. These samples enabled direct comparisons between



the DNA methylation profile and homoeolog expression patterns in the same samples. WGBS libraries were constructed from purified nuclei prepared using the published methods (51). Input DNA was quantified using the Qubit high sensitivity DNA kit. A total of 500 ng of nuclear DNA was spiked with 270 pg of Lambda DNA to assess the conversion efficiency obtained using the EZ DNA Methylation-GoldTM Kit (Zymo research corp, Irvine, Ca, USA). WGBS libraries were prepared using the TruSeq DNA kit, (Illumina, Madison, WI) and 2 × 125 bp paired-end sequence reads was generated using the Illumina HiSeq 2500 v4 platform (Genome Quebec, Montreal, QC, Canada). The data was deposited as SRP133674.

### DNA methylation data analysis

Sequence quality and adaptor removal was performed using Trim\_galore\_v0.4.1 (52). High quality paired-end sequence reads were aligned to the RefSeqv1.0 Chinese Spring genome using Bismark version 0.16.1 (53) ensuring the removal of duplicate reads and only retaining unique unambiguous alignments. The data were processed to exclude regions with low coverage using a binomial test. The methylation data was annotated using the gene feature coordinates provided by the RefSeqv1.0 Chinese Spring gene definitions. 5 kb flanking regions around the gene features were also extracted. The two flanking regions and the gene feature were each divided into 50 tiles (150 tiles in total) to summarize the observed methylation ratios. Data manipulation, statistical analysis and image generation were performed using the R language (54) utilizing the data.table (55), MethylKit v1.5.2 (56), genomation (57), and ggplot2 packages (58).

### Comparison of RNA-seq sample classification with DNA methylation

For the five RNA-seq samples (from the same plants used for analyzing DNA methylation) we classified triads into the seven balanced, dominant, and suppressed categories using the same method as for previous analyses. We then classified homoeologs within dominant and suppressed triads into the “dominant” and “nondominant” homoeologs, and “suppressed” and “nonsuppressed” homoeologs. For example, in an A dominant triad the A subgenome is classified as “dominant” and B and D subgenomes are classified as “nondominant” homoeologs. The DNA methylation patterns of genes in each of these categories were plotted using the methods described above (DNA methylation data analysis). Differences in DNA methylation levels between categories were tested pairwise using the non-parametric Mann-Whitney *t* test using the wilcox.test() in R (fig. S17).

### Histone modification analysis

To study the role of histone modifications we carried out ChIP-seq for three active marks (H3K36me3, H3K9ac, and H3K4me3) and one repressive mark (H3K27me3) (deposited under SRA accession number SRP126222). We used

Chinese Spring at 3-leaf stage, however RNA-seq data were not collected from these exact plants. To calculate the homoeolog expression bias we used Chinese Spring samples from a separate experiment (*PAMP-triggered immune response* study) in which the same tissue was collected at a similar stage (3-leaf stage). Whilst combining data from two separate experiments may introduce some noise into the analysis, the ChIP-seq and RNA-seq data are from similar tissues, at a similar growth stage, in the same wheat variety, and are thus highly comparable. Nevertheless, this confounding factor should be considered when interpreting these results. ChIP assays, DNA library preparation, and sequencing were performed as in (10).

### Histone data analysis

Raw FASTQ files were preprocessed with Trimmomatic v0.36 (59) to remove Illumina sequencing adapters, trim 5' and 3' ends with quality score below 5 (Phred+33) and discard reads shorter than 20 bp after trimming. Paired-ends reads were aligned against IWGSC RefSeq v1.0 assembly using bowtie2 v2.3.3 with -very-sensitive settings (60). Alignments with MAPQ < 10 were discarded and duplicate reads removed with Picard MarkDuplicates (<http://broadinstitute.github.io/picard/>). Triad expression category was calculated using Chinese Spring samples from a separate experiment (*PAMP-triggered immune response* study) using the same method as in previous analyses. As in the DNA methylation analysis we classified homoeologs within dominant and suppressed triads into the “dominant” and “nondominant” homoeologs, and “suppressed” and “nonsuppressed” homoeologs.

We calculated meta-gene profiles for each category by computing the read density of each histone mark over different triads categories using Deeptools (61) *computeMatrix scale-regions* and plotted it with *plotProfile*. To make a statistical comparison, for each histone mark we scored the number of reads overlapping with gene bodies using *bedtools coverage -counts* (62). Only reads fully mapping within gene bodies were considered. To account for different gene size we divided the read counts over each gene by its length. The distributions of reads density over different triads categories were compared with a nonparametric *t* test (Mann-Whitney U-test) using the function *wilcox.test* in R (fig. S11).

### Variation in homoeolog expression bias across tissues (stable and dynamic triads)

To define the variation in homoeolog expression bias of each triad across the intermediate tissues we calculated the Euclidean distance between the triad's global position (*combined analysis*) and each individual tissue in which the triad was considered expressed. We included only triads which were considered expressed in at least six tissues based on the *combined analysis* criteria outlined above. The average of these distances was defined as the “*triad mean distance*”. We

ranked triads by their *triad mean distance* and the percentile was calculated by

$$\text{percentile}_i = \text{truncate} \left( \frac{\text{rank}(\text{cmd}_i)}{\text{length}(\text{CMD})} \right)$$

where CMD is the vector containing all the *triad mean distance*. The first and last deciles were classified as stable 10% and dynamic 10% triads, respectively. A similar approach was used to define the corresponding 5% and 25% extremes of the distribution. This analysis was conducted independently for the Chinese Spring no stress samples, the Azhurnaya developmental time course, and for each of the four tissue-specific networks. A visual representation is provided in fig. S19.

### TE presence in gene promoters

We extracted all TEs that were annotated to fall at least partly within 1.5 kb and 5 kb upstream of the canonical ATG start-codon for all genes. We then split the TEs into the relevant gene lists covering homoeolog expression bias variation (stable 10%, middle 80%, and dynamic 10%) and homoeolog expression bias (balanced, dominant, nondominant, suppressed, and nonsuppressed) based on the “combined analysis (all tissues)” for CS. We used these lists to identify the proportion of genes and triads in each category which contained at least one TE in the promoter region.

### Enrichment of TE families in gene promoters

Using the GFF file of TE coordinates, we extracted TEs present in the promoter regions of HC genes. We retrieved all TE copies that are entirely or partially present in the 5 kb upstream of the ATG start-codon of the canonical transcript for each gene. We then calculated the number of genes in each of the stable 10%, middle 80%, and dynamic 10% categories which contained specific TE families. We required the TE family to be present in at least 2% of the categorized genes for further analysis. We then found the deviation of this distribution from the expected 10-80-10 ratio using the  $\chi^2$  test, *P* values adjusted with Benjamini-Hochberg. We calculated the median length of each TE family based on all instances of that TE across the genome. We found fifteen TE families deviated significantly from the expected 10-80-10 distribution (Benjamini-Hochberg *P* < 0.01). However, the majority of these TE families were present in less than 5% of the genes considered, and showed very small variation in the number of promoters containing the TE, suggesting that the statistical significance may not be biologically relevant.

### TE density in gene promoters

We calculated the density of TEs within 5 kb upstream of genes by calculating the proportion of TE bases in sliding windows of 100 bp with a step size of 10 bp. The mean of each window was then calculated for both the stable 10%, middle 80%, and dynamic 10% triads and

the subgenome dominance categories (balanced, dominant, nondominant, suppressed, and non-suppressed). Mann Whitney tests with Benjamini-Hochberg adjusted *P* values were used to test for differences in TE density between categories across each window.

### Transcription factor binding site identification

The 1.5 kb of sequence upstream of the canonical ATG start-codon was used to identify transcription factor binding sites (TFBS) present in promoters of HC triads. The FIMO tool from the MEME suite [v 4.11.4 (63)] was used with a position weight matrix (PWM) obtained from plantPAN 2.0 (64) to predict TFBS based on previously identified sites across multiple plant species. FIMO was run with a *P* value threshold of  $<1\text{E-}04$  (default), -motif-pseudo set to  $1\text{E-}08$  as recommended for use with PWMs and a -max-stored-scores of 1,000,000 to account for the large size of the dataset. The background model was generated from all extracted promoter sequences using the MEME fasta-get-markov command. Details of the TFBS comparisons between homoeologs is presented in the supplementary materials (13).

### WGCNA network construction

Coexpression networks were built for six separate sample sets: grain, leaf, spike, root, abiotic and disease (table S1) using the WGCNA R package (65). For each network, we selected HC genes which were expressed  $> 0.5$  TPM in three or more samples. The count expression level of each gene was normalized using variance stabilizing transformation from DESeq2 (66) to eliminate differences in sequencing depth between studies. The soft power threshold was calculated as the first power to exceed a scale-free topology fit index of 0.9 for each network separately. The soft powers used were: leaf = 12, spike = 12, roots = 7, disease = 7. For the abiotic and grain network the 0.9 threshold was not crossed until 15 and 20 respectively, which may be due to strong differences between samples within these datasets, therefore the soft power threshold was selected according to the number of samples, resulting in abiotic = 7 and grain = 6. Signed hybrid networks were constructed blockwise using the function blockwiseModules() with a maximum block size of 46,000 genes. The correlation type used was biweight mid-correlation “bicor” and the maxPOutliers was set to 0.05 to eliminate effects of outlier samples. The topographical overlap matrices (TOM) were calculated by the blockwiseModules function using TOMType = “unsigned” and the minimum module size was set to 30. The parameter mergeCutHeight = 0.15 was used to merge similar modules.

### Defining same, similar, and divergent expression patterns of triads

For triads which had homoeologs within different modules in the WGCNA networks we developed a threshold to determine whether the different modules had a “similar” or “di-

vergent” expression pattern. We calculated the Euclidean distance between module eigengenes using the R package dist() and with these values we calculate the distances between the homoeologs in each triad. Triads where the pairwise distances were zero were in the same module. Triads where the pairwise distances were over zero were in different modules. For these triads in different modules when the pairwise distance between any two homoeologs was  $> 50\%$  of the median maximum distance between eigengenes, the triad was classified as having a “divergent” expression pattern. In cases where the pairwise distance was over zero between at least one pair of homoeologs and the distance between all three pairs of homoeologs were  $\leq 50\%$  of the median maximum distance, the triad was classified as having a “similar” expression pattern. The median maximum distance between eigengenes was averaged across all four tissue networks to give a final threshold (50% of median maximum distance) of 0.937431. This analysis was carried out for 1:1:1 syntenic and 1:1:1 nonsyntenic triads expressed in each of the tissue networks (total triads = 9599 grain, 5378 leaf, 11,038 root, and 6173 spike). This excluded triads which had a putative transposable element (67 triads). Fig. S23 shows a graphic representation of this classification and the effect of altering the threshold in each of the four networks.

### Module overlaps

Module overlap between networks was calculated using the R package GeneOverlap which calculates significant overlaps between modules using a Fisher's exact test. Modules were considered to have significant overlaps when the FDR adjusted *P* value  $< 0.05$ .

### Correlation to stress status

Modules within the abiotic and disease networks were tested for correlations to intermediate level stresses using the cor(function). The significant of correlations were calculated using the function corPvalueStudent() and corrected for multiple testing using p.adjust() using the Benjamini & Yekutieli method (67).

### Genie3

HC genes expressed  $>5$  TPM in at least one of the 850 sample were selected. Out of these 78,085 genes there were 3386 transcription factors (methods described above). Random forest regression was estimated for each gene based on the transcription factors as inputs using the genie3 package (68) in R (version 3.3.2) with default parameters (K=sqrt, nb.trees=1000, input.idx=list of transcription factors, importance.measure=IncNodePurity, seed=NULL). For each transcription factor, all predicted target genes (connectivity  $> 0.005$ ) were extracted and functional enrichment within the target genes was determined using topGO (69) in R (version 3.3.2) with the following parameters (ontology = “BP”, nodeSize = 10, classic Fisher test  $P < 10^{-10}$ ). To summarize the results, the top three GO terms for each transcription factor, the *P* values for the

strongest enrichment, and the direct blastx match in the *Arabidopsis* proteome (tair10) and well as the e-value and description were tabulated (table S28). The complete list of GO term enrichments of the biological process ontology for each transcription factor and the list of transcription factors associated with each GO term in the ontology of biological process are published in e!DAL (<https://doi.ipk-gatersleben.de/DOI/53148abd-26a1-4ede-802b-c2635aff6a725/0dd8224a-34fc-4e3b-8ab8-883d07e52bd2/2/1847940088>).

### Identifying highly connected hub genes

Hub genes within each module for the abiotic and disease stress networks were calculated using the WGCNA R package function signedKME(). This calculates the correlation between the expression patterns of each gene and the module eigengene. Genes which were more highly correlated to the eigengene were considered hub genes.

### REFERENCES AND NOTES

- W. Albertin, P. Marullo, Polyploidy in fungi: Evolution after whole-genome duplication. *Proc. Biol. Sci.* **279**, 2497–2509 (2012). doi: [10.1098/rspb.2012.0434](https://doi.org/10.1098/rspb.2012.0434); pmid: [22492065](https://pubmed.ncbi.nlm.nih.gov/22492065/)
- S. P. Otto, J. Whitton, Polyploid incidence and evolution. *Annu. Rev. Genet.* **34**, 401–437 (2000). doi: [10.1146/annurev.genet.34.1.401](https://doi.org/10.1146/annurev.genet.34.1.401); pmid: [11092833](https://pubmed.ncbi.nlm.nih.gov/11092833/)
- A. Salman-Minkov, N. Sabath, I. Mayrose, Whole-genome duplication as a key factor in crop domestication. *Nat. Plants* **2**, 16115 (2016). doi: [10.1038/nplants.2016.115](https://doi.org/10.1038/nplants.2016.115); pmid: [27479829](https://pubmed.ncbi.nlm.nih.gov/27479829/)
- S. Renny-Byfield, J. F. Wendel, Doubling down on genomes: Polyploidy and crop plants. *Am. J. Bot.* **101**, 1711–1725 (2014). doi: [10.3732/ajb.1400119](https://doi.org/10.3732/ajb.1400119); pmid: [25090999](https://pubmed.ncbi.nlm.nih.gov/25090999/)
- M. Feldman, A. A. Levy, T. Fahima, A. Korol, Genomic asymmetry in allopolyploid plants: Wheat as a model. *J. Exp. Bot.* **63**, 5045–5059 (2012). doi: [10.1093/jxb/ers192](https://doi.org/10.1093/jxb/ers192); pmid: [22859676](https://pubmed.ncbi.nlm.nih.gov/22859676/)
- Y. Van de Peer, E. Mizrahi, K. Marchal, The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **18**, 411–424 (2017). doi: [10.1038/nrg.2017.26](https://doi.org/10.1038/nrg.2017.26); pmid: [28502977](https://pubmed.ncbi.nlm.nih.gov/28502977/)
- S. Ohno, *Evolution by Gene Duplication* (Springer-Verlag, 1970).
- F. A. Kondrashov, Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc. Biol. Sci.* **279**, 5048–5057 (2012). doi: [10.1098/rspb.2012.1108](https://doi.org/10.1098/rspb.2012.1108); pmid: [22977152](https://pubmed.ncbi.nlm.nih.gov/22977152/)
- K. D. Makova, W.-H. Li, Divergence in the spatial pattern of gene expression between human duplicate genes. *Genome Res.* **13**, 1638–1645 (2003). doi: [10.1101/gr.1133803](https://doi.org/10.1101/gr.1133803); pmid: [12840042](https://pubmed.ncbi.nlm.nih.gov/12840042/)
- International Wheat Genome Sequencing Consortium, Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361**, eaar7191 (2018). doi: [10.1126/science.aar7191](https://doi.org/10.1126/science.aar7191)
- E. Martinez-Perez, P. Shaw, G. Moore, The *Ph1* locus is needed to ensure specific somatic and meiotic centromere association. *Nature* **411**, 204–207 (2001). doi: [10.1038/35075597](https://doi.org/10.1038/35075597); pmid: [11346798](https://pubmed.ncbi.nlm.nih.gov/11346798/)
- G. Moore, K. M. Devos, Z. Wang, M. D. Gale, Cereal genome evolution. Grasses, line up and form a circle. *Curr. Biol.* **5**, 737–739 (1995). doi: [10.1016/S0960-9822\(95\)00148-5](https://doi.org/10.1016/S0960-9822(95)00148-5); pmid: [7583118](https://pubmed.ncbi.nlm.nih.gov/7583118/)
- Additional materials and methods are available as supplementary materials.
- N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016). doi: [10.1038/nbt.3519](https://doi.org/10.1038/nbt.3519); pmid: [27043002](https://pubmed.ncbi.nlm.nih.gov/27043002/)
- P. Borrill, R. Ramirez-Gonzalez, C. Uauy, expVIP: A customisable RNA-seq data analysis and visualization platform. *Plant Physiol.* **170**, 2172–2186 (2016). doi: [10.1104/pp.15.01667](https://doi.org/10.1104/pp.15.01667); pmid: [26869702](https://pubmed.ncbi.nlm.nih.gov/26869702/)
- M. Melé et al., The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015). doi: [10.1126/science.aaa0355](https://doi.org/10.1126/science.aaa0355); pmid: [25954002](https://pubmed.ncbi.nlm.nih.gov/25954002/)
- J. W. Walley et al., Integration of omic networks in a developmental atlas of maize. *Science* **353**, 814–818 (2016). doi: [10.1126/science.aag1125](https://doi.org/10.1126/science.aag1125); pmid: [27540173](https://pubmed.ncbi.nlm.nih.gov/27540173/)



18. M. Pfeifer *et al.*, Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science* **345**, 1250091 (2014). doi: [10.1126/science.1250091](https://doi.org/10.1126/science.1250091); pmid: 25035498
19. D. Winter *et al.*, An “Electronic Fluorescent Pictograph” browser for exploring and analyzing large-scale biological data sets. *PLOS ONE* **2**, e718 (2007). doi: [10.1371/journal.pone.0000718](https://doi.org/10.1371/journal.pone.0000718); pmid: 17684564
20. P. Borrill, N. Adamski, C. Uauy, Genomics as the key to unlocking the polyploid potential of wheat. *New Phytol.* **208**, 1008–1022 (2015). doi: [10.1111/nph.13533](https://doi.org/10.1111/nph.13533); pmid: 26108556
21. R. Avni *et al.*, Functional characterization of GPC-1 genes in hexaploid wheat. *Planta* **239**, 313–324 (2014). doi: [10.1007/s00425-013-1977-y](https://doi.org/10.1007/s00425-013-1977-y); pmid: 24170335
22. K. J. Simons *et al.*, Molecular characterization of the major wheat domestication gene *Q. Genetics* **172**, 547–555 (2006). doi: [10.1534/genetics.105.044727](https://doi.org/10.1534/genetics.105.044727); pmid: 16172507
23. C. E. Grover *et al.*, Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytol.* **196**, 966–971 (2012). doi: [10.1111/j.1469-8137.2012.04365.x](https://doi.org/10.1111/j.1469-8137.2012.04365.x); pmid: 23033870
24. S. Renny-Byfield *et al.*, Ancient gene duplicates in *Gossypium* (cotton) exhibit near-complete expression divergence. *Genome Biol. Evol.* **6**, 559–571 (2014). doi: [10.1093/gbe/evu037](https://doi.org/10.1093/gbe/evu037); pmid: 24558256
25. M. Hao *et al.*, The abundance of homoeologue transcripts is disrupted by hybridization and is partially restored by genome doubling in synthetic hexaploid wheat. *BMC Genomics* **18**, 149 (2017). doi: [10.1186/s12864-017-3558-0](https://doi.org/10.1186/s12864-017-3558-0); pmid: 28187716
26. M. J. Yoo, E. Szadkowski, J. F. Wendel, Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* **110**, 171–180 (2013). doi: [10.1038/hdy.2012.94](https://doi.org/10.1038/hdy.2012.94); pmid: 23169565
27. P. P. Edger *et al.*, Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *Plant Cell* **29**, 2150–2167 (2017). doi: [10.1105/tpc.17.00010](https://doi.org/10.1105/tpc.17.00010); pmid: 28814644
28. A. J. Bewick, R. J. Schmitz, Gene body DNA methylation in plants. *Curr. Opin. Plant Biol.* **36**, 103–110 (2017). doi: [10.1016/j.pbi.2016.12.007](https://doi.org/10.1016/j.pbi.2016.12.007); pmid: 28258985
29. D. Zilberman, An evolutionary case for functional gene body methylation in plants and animals. *Genome Biol.* **18**, 87 (2017). doi: [10.1186/s13059-017-1230-2](https://doi.org/10.1186/s13059-017-1230-2); pmid: 28486944
30. A. J. Bewick *et al.*, On the origin and evolutionary consequences of gene body DNA methylation. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 9111–9116 (2016). doi: [10.1073/pnas.1604666113](https://doi.org/10.1073/pnas.1604666113); pmid: 27457936
31. X. Zhang *et al.*, Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. *PLOS Biol.* **5**, e129 (2007). doi: [10.1371/journal.pbio.0050129](https://doi.org/10.1371/journal.pbio.0050129); pmid: 17439305
32. E. D. Akhunov *et al.*, The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. *Genome Res.* **13**, 753–763 (2003). doi: [10.1101/gr.080603](https://doi.org/10.1101/gr.080603); pmid: 12695326
33. R. J. A. Bugge *et al.*, Tissue-specific silencing of homoeologs in natural populations of the recent allopolyploid *Tragopogon mirus*. *New Phytol.* **186**, 175–183 (2010). doi: [10.1111/j.1469-8137.2010.03205.x](https://doi.org/10.1111/j.1469-8137.2010.03205.x); pmid: 20409177
34. H. Zhao *et al.*, Proliferation of regulatory DNA elements derived from transposable elements in the maize genome. *Plant Physiol.* **176**, 2789–2803 (2018). doi: [10.1104/pp.17.01467](https://doi.org/10.1104/pp.17.01467); pmid: 29463772
35. C. D. Hirsch, N. M. Springer, Transposable element influences on gene expression in plants. *BBA Gene Regul. Mech.* **1860**, 157–165 (2017).
36. A. Roulin *et al.*, The fate of duplicated genes in a polyploid plant genome. *Plant J.* **73**, 143–153 (2013). doi: [10.1111/tbj.12026](https://doi.org/10.1111/tbj.12026); pmid: 22974547
37. J.-T. Li *et al.*, The fate of recent duplicated genes following a fourth-round whole genome duplication in a tetraploid fish, common carp (*Cyprinus carpio*). *Sci. Rep.* **5**, 8199 (2015). doi: [10.1038/srep08199](https://doi.org/10.1038/srep08199); pmid: 25645996
38. C. Yu *et al.*, MADS-box transcription factor *OsMADS25* regulates root development through affection of nitrate accumulation in rice. *PLOS ONE* **10**, e0135196 (2015). doi: [10.1371/journal.pone.0135196](https://doi.org/10.1371/journal.pone.0135196); pmid: 26258667
39. L.-H. Yu *et al.*, MADS-box transcription factor *AGL21* regulates lateral root development and responds to multiple external and physiological signals. *Mol. Plant J.* **7**, 1653–1669 (2014). doi: [10.1093/mp/ssu088](https://doi.org/10.1093/mp/ssu088); pmid: 25122697
40. S. Guo *et al.*, The interaction between *OsMADS57* and *OSTB1* modulates rice tillering via *DWARF14*. *Nat. Commun.* **4**, 1566 (2013). doi: [10.1038/ncomms2542](https://doi.org/10.1038/ncomms2542); pmid: 23463009
41. A. Ghazalpour *et al.*, Integrating genetic and network analysis to characterize genes related to mouse weight. *PLOS Genet.* **2**, e130 (2006). doi: [10.1371/journal.pgen.0020130](https://doi.org/10.1371/journal.pgen.0020130); pmid: 16934000
42. M. Kumar *et al.*, Heat shock factors HsfB1 and HsfB2b are involved in the regulation of *Pdfl2* expression and pathogen resistance in *Arabidopsis*. *Mol. Plant* **2**, 152–165 (2009). doi: [10.1093/mp/ssn095](https://doi.org/10.1093/mp/ssn095); pmid: 19529832
43. K. M. Pajerowska-Mukhtar *et al.*, The HSF-like transcription factor TBF1 is a major molecular switch for plant growth-to-defense transition. *Curr. Biol.* **22**, 103–112 (2012). doi: [10.1016/j.cub.2011.12.015](https://doi.org/10.1016/j.cub.2011.12.015); pmid: 22244999
44. M. Ikeda, N. Mitsuda, M. Ohme-Takagi, Arabidopsis HsfB1 and HsfB2b act as repressors of the expression of heat-inducible Hsfs but positively regulate the acquired thermotolerance. *Plant Physiol.* **157**, 1243–1254 (2011). doi: [10.1104/pp.111.179036](https://doi.org/10.1104/pp.111.179036); pmid: 21908690
45. G. Xu *et al.*, uORF-mediated translation allows engineered plant disease resistance without fitness costs. *Nature* **545**, 491–494 (2017). doi: [10.1038/nature22372](https://doi.org/10.1038/nature22372); pmid: 28514448
46. K. V. Krasileva *et al.*, Uncovering hidden variation in polyploid wheat. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E913–E921 (2017). doi: [10.1073/pnas.1619268114](https://doi.org/10.1073/pnas.1619268114); pmid: 28096351
47. Y. Zhang *et al.*, Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nat. Commun.* **7**, 12617 (2016). doi: [10.1038/ncomms12617](https://doi.org/10.1038/ncomms12617); pmid: 27558837
48. FAO, [www.fao.org/faostat/](http://www.fao.org/faostat/).
49. L. J. Leach *et al.*, Patterns of homoeologous gene expression shown by RNA sequencing in hexaploid bread wheat. *BMC Genomics* **15**, 276 (2014). doi: [10.1186/1471-2164-15-276](https://doi.org/10.1186/1471-2164-15-276); pmid: 24726045
50. N. Hamilton, ggtern: An extension to ‘ggplot2’, for the creation of ternary diagrams. R package version 2.2.1 (2016); <https://CRAN.R-project.org/package=ggtern>.
51. H.-B. Zhang, X. Zhao, X. Ding, A. H. Paterson, R. A. Wing, Preparation of megabase-size DNA from plant nuclei. *Plant J.* **7**, 175–184 (1995). doi: [10.1046/j.1365-3113.1995.07010175.x](https://doi.org/10.1046/j.1365-3113.1995.07010175.x)
52. S. Lindgreen, AdapterRemoval: Easy cleaning of next-generation sequencing reads. *BMC Res. Notes* **5**, 337 (2012). doi: [10.1186/1756-0500-5-337](https://doi.org/10.1186/1756-0500-5-337); pmid: 22748135
53. F. Krueger, S. R. Andrews, Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011). doi: [10.1093/bioinformatics/btr167](https://doi.org/10.1093/bioinformatics/btr167); pmid: 21493656
54. R Core Team, R: A language and environment for statistical computing (R Foundation for Statistical Computing, 2013); [www.R-project.org/](http://www.R-project.org/).
55. M. Dowle, M. Srinivasan, data.table: Extension of ‘data.frame’. R package version 1.10.4-3 (2017); <https://CRAN.R-project.org/package=data.table>.
56. A. Akalin *et al.*, methylKit: A comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13**, R87 (2012). doi: [10.1186/gb-2012-13-10-R87](https://doi.org/10.1186/gb-2012-13-10-R87); pmid: 23034086
57. A. Akalin, V. Franke, K. Vlahovicek, C. E. Mason, D. Schübeler, Genomation: A toolkit to summarize, annotate and visualize genomic intervals. *Bioinformatics* **31**, 1127–1129 (2015). doi: [10.1093/bioinformatics/btu775](https://doi.org/10.1093/bioinformatics/btu775); pmid: 25417204
58. H. Wickham, ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag, 2009).
59. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014). doi: [10.1093/bioinformatics/btu170](https://doi.org/10.1093/bioinformatics/btu170); pmid: 24695404
60. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012). doi: [10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923); pmid: 22388286
61. F. Ramírez *et al.*, deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016). doi: [10.1093/nar/gkw257](https://doi.org/10.1093/nar/gkw257); pmid: 27079975
62. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010). doi: [10.1093/bioinformatics/btq033](https://doi.org/10.1093/bioinformatics/btq033); pmid: 20110278
63. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011). doi: [10.1093/bioinformatics/btr064](https://doi.org/10.1093/bioinformatics/btr064); pmid: 21330290
64. C.-N. Chow *et al.*, PlantPAN 2.0: An update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. *Nucleic Acids Res.* **44**, D1154–D1160 (2016). doi: [10.1093/nar/gkv1035](https://doi.org/10.1093/nar/gkv1035); pmid: 26476450
65. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008). doi: [10.1186/1471-2105-9-559](https://doi.org/10.1186/1471-2105-9-559); pmid: 19114008
66. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014). doi: [10.1186/s13059-014-0550-8](https://doi.org/10.1186/s13059-014-0550-8); pmid: 25516281
67. Y. Benjamini, D. Yekutieli, The Control of the False Discovery Rate in Multiple Testing under Dependency. *Ann. Stat.* **29**, 1165–1188 (2001).
68. V. A. Huynh-Thu, A. Irrthum, L. Wehenkel, P. Geurts, Inferring regulatory networks from expression data using tree-based methods. *PLOS ONE* **5**, e12776 (2010). doi: [10.1371/journal.pone.0012776](https://doi.org/10.1371/journal.pone.0012776); pmid: 20927193
69. A. Alexa, J. Rahnenfuhrer, topGO: Enrichment analysis for gene ontology. R package version 2.30.0 (2016). doi: [10.18129/B9.bioc.topGO](https://doi.org/10.18129/B9.bioc.topGO)

## ACKNOWLEDGMENTS

We thank Bayer Crop Science staff members E. Caestecker and X. Wang for data analysis and B. Staelens, T. Debaecke, and A. Dobbelaere for plant growth and sampling. We also acknowledge the assistance of M. Burrell (NBI Computing) and A. Etuk (EI Digital Biology). **Funding:** This work was supported by the UK Biotechnology and Biological Sciences Research Council (BBSRC) through the Designing Future Wheat (BB/P016855/1), GEN (BB/P013511/1), and Plant Health (BB/P012574/1) ISPs; Tritecrae Genomics for Sustainable Agriculture (BB/J003557/1); ERA-PG (BB/G024960/1); ERA-CAPS (BB/N005007/1); and an Anniversary Future Leaders Fellowship to P.B. (BB/M014045/1). This work was also supported by the International Wheat Yield Partnership (IWYP76); the German Federal Ministry of Food and Agriculture (281913915); the German Ministry of Education and Research (031A536); DFG (SFB924); Genome Canada/Ontario Genomics (OGI-128); the Canadian Applied Triticum Genomics project (CTAG2) funded by Genome Canada, Genome Prairie, Western Grains Research Foundation, Saskatchewan Wheat Development Commission, Alberta Wheat Development Commission, and Saskatchewan Ministry of Agriculture; the National Research Council of Canada Wheat Flagship program; the WGRCI/UCRC partially funded by NSF (IIP-1338897); French Agence Nationale de la Recherche grants ANR-11-BSV5-0015 and ANR-16-TERC-0026-01; and the European Research Council. S.A.H. was supported by the John Innes Foundation. C.J. was supported by Région Auvergne and the European Regional Development Fund (SRESRI 2016). This research was also supported in part by the NBI Computing infrastructure for Science (CiS) group through the HPC resources. The submission of sequencing data was brokered by the COPO platform (<https://copo-project.org>), funded by the BBSRC (BB/L024055/1), and supported by CyVerse UK, part of the Earlham Institute National Capability in e-Infrastructure. **Author contributions:** R.H.R.-G., P.B., and C.U. conceived, designed, and coordinated the study. R.H.R.-G. and P.B. organized RNA-seq samples and assigned metadata. R.H.R.-G. carried out the mapping and developed methods to analyze homoeolog expression patterns. R.H.R.-G. and C.U. analyzed homoeolog expression patterns and variation between tissues, cultivars, syntenic and nonsyntenic triads, chromosomal partitions, and progenitor species. P.B. constructed WGCNA coexpression networks, analyzed homoeolog coexpression, and identified biological case studies in the networks. P.B. carried out differential expression analysis between tissues. D.Lan., K.F.X.M., and M.S. generated gene annotations, performed phylogenomics analysis (including gene family, ortholog and homoeolog inference, phylogenetic trees, and TF superfamily classification), and analyzed subgenome expression bias. A.P. led informatics development, and Y.K. generated pictograph drawings for the wheat eFP portal, supervised by A.G.S. and N.J.P. A.B.R. ran and analyzed the genie3 network. S.A.H. analyzed  $K_A/K_S$  ratios. A.T.C., S.J.R., and A.G.S. performed and analyzed DNA methylation profiles. D.Lat. performed ChIP-seq experiments, and L.C. performed bioinformatic analysis of ChIP-seq data, supervised by M.B. and A.Be. E.P. and C.J. analyzed histone marks. S.A.H., J.B., R.H.R.-G., and L.V. carried out the promoter analysis. T.F. illustrated wheat development for Fig. 1. E.P. identified chromosome partitions. F.C. identified and S.A.H. and J.B. analyzed transposable elements. J.B. conducted promoter cis-regulatory element analysis. M.D., J.J., F.V.E., S.J.R., A.T.C., H.S., B.S., D.X., C.J.R., B.C., B.B.H.W., R.A., V.T., R.D., C.J.P., and A.G.S. provided RNA-seq samples. P.B. and C.U. wrote the manuscript. R.H.R.-G., S.A.H., J.B., A.B.R., D.Lan., A.T.C., L.C., C.J., E.P., and T.F. contributed text and figures for the manuscript. L.V., M.D., J.J., F.V.E., Y.K., J.B., H.S., B.S., C.J.R., R.A., V.T., R.D., F.C., C.J.P., N.J.P.,

A.G.S., and M.S. provided comments on the manuscript. All authors read and approved the final submission. **Competing interests:** The authors declare no conflicts of interest. **Data and materials availability:** All code used for the analyses of the datasets can be found at <https://github.com/Uauy-Lab/WheatHomoeologExpression>, and data files are deposited in <https://grassroots.tools>. Sequencing reads were deposited with NCBI under accession codes PRJEB25639, PRJEB23056, PRJNA436817, SRP133837, PRJEB25640, and PRJEB25593 for RNA-seq; SRP133674

for whole-genome bisulfite sequencing; and SRP126222 for ChIP-seq. The full phylogenetic tree presented in Fig. 4C is available at <http://itol.embl.de/shared/borriillp>. Details of each dataset can be found in the methods and the supplementary materials.

#### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/361/6403/eaar6089/suppl/DC1](http://www.sciencemag.org/content/361/6403/eaar6089/suppl/DC1)  
Additional Materials and Methods

Figs. S1 to S24  
Tables S1 to S37  
IWGSC Collaborator List  
References (70–97)  
Data S1 and S2

30 November 2017; accepted 11 July 2018  
10.1126/science.aar6089



## RESEARCH ARTICLE SUMMARY

## DEVELOPMENTAL SIGNALING

# A molecular mechanism for Wnt ligand-specific signaling

Marie Eubelen\*, Naguissa Bostaille\*, Pauline Cabochette, Anne Gauquier, Patricia Tebabi, Andra C. Dumitru, Melanie Koehler, Philipp Gut, David Alsteens, Didier Y. R. Stainier, Abel Garcia-Pino, Benoit Vanhollebeke†

**INTRODUCTION:** Wnt signaling is an ancient signaling pathway that has accompanied the emergence of metazoans and is key to many developmental, physiological, and disease processes. Similar to other signaling pathways, gene families for both Wnt ligand and its corresponding Frizzled receptor have undergone extensive expansion during metazoan evolution. Vertebrate genomes harbor 19 closely related Wnt genes as well as 10 Frizzled genes. Gene duplication is typically considered a major driving force in the evolution of new biological functions through neo- or subfunctionalization of emerging paralogs. How this functional diversification of Wnt ligands is structurally and molecularly organized, however, remains poorly understood. The Wnt/Frizzled molecular interaction is mediated by residues conserved across

both families. This promiscuous interaction is incompatible with monospecific recognition and, accordingly, when tested in pair-wise combinations, multiple Wnt ligands compete for binding to various Frizzled receptors.

**RATIONALE:** These observations raise the questions of how Wnt ligands achieve functional diversification and how cells interpret the intermingled expression patterns of simultaneous and sometimes conflicting Wnt signals. In some biological settings, cells may integrate all signaling inputs nondiscriminately and trigger appropriate responses by considering their total net balance. However, other biological processes exhibit strict Wnt ligand selectivity, despite complex Wnt/Frizzled expression landscapes. A prototypical example

is provided by the exclusive control of mammalian forebrain and ventral spinal cord angiogenesis by Wnt7a and Wnt7b.

Within this neurovascular unit, in order to respond to neural progenitor-derived Wnt7 by activating Wnt/ $\beta$ -catenin signaling, cerebral endothelial cells must express a membrane protein complex consisting of the adhesion G protein-coupled receptor (GPCR) Gpr124 and the glycosylphosphatidylinositol-anchored glycoprotein Reck. This Gpr124/Reck complex was recently reported to promote Wnt7-specific responses.

**RESULTS:** Using a combination of biophysical approaches and ligand-binding assays in genetically engineered cells, we demonstrate that ligand selectivity is conferred by Reck, which mediates Wnt7-specific binding in a Frizzled-

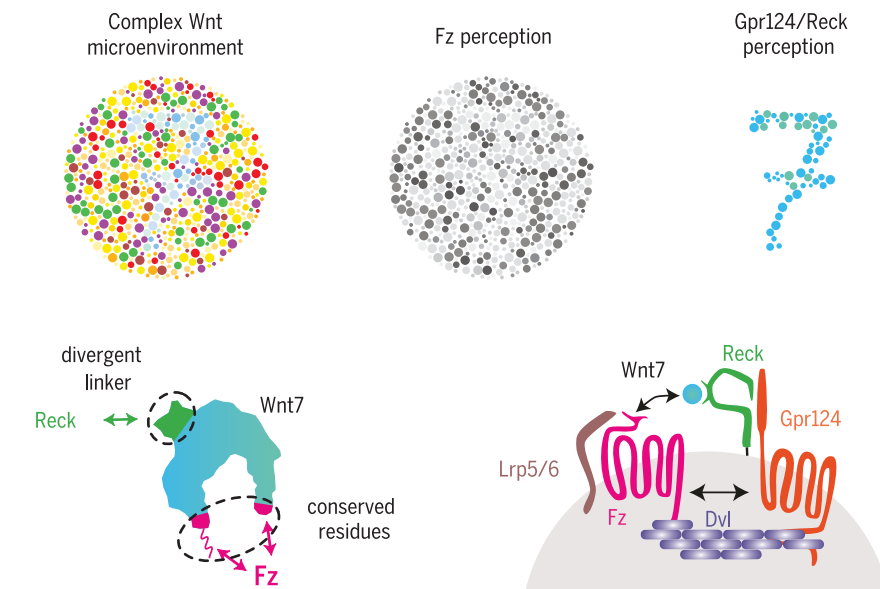
## ON OUR WEBSITE

Read the full article at <http://dx.doi.org/10.1126/science.aat1178>

independent manner. Reck orchestrates Wnt ligand discrimination by engaging the structurally disordered and highly divergent linker domain of Wnt7. The presence of Gpr124 is required

to deliver Reck-bound Wnt7 to Frizzled by assembling higher-order Reck/Gpr124/Frizzled/Lrp5/6 complexes. This Gpr124 tethering function does not rely on its GPCR structure but instead on its combined capacity to interact with Reck extracellularly and recruit the Dishevelled scaffolding protein intracellularly. By bridging Gpr124 and Frizzled, Dishevelled recruits Wnt7, via its association with Reck, into dynamic Wnt/Frizzled signalosomes, resulting in increased local concentrations of ligand available for Frizzled signaling.

**CONCLUSION:** Our data reveal that cells are equipped with “Wnt-decoding modules” that distinguish between Wnt ligands that are otherwise very similar. They also reveal a critical role for the linker domain in Wnt ligand evolution and functional diversification. These mechanistic insights into the Wnt decoding capacities of vertebrate cells predict that additional Wnt decoding modules exist, enabling fine-tuning of cellular behaviors in response to other Wnt or Frizzled family members. These modules expand the diversity of proximal events in Wnt signaling, opening new therapeutic opportunities for conditions in which Wnt stimulation or inhibition are desirable at the membrane level. In particular, the mechanisms uncovered here provide an opportunity for the targeted treatment of human central nervous system neurovascular disorders. ■



**Task sharing for orchestrated Wnt7-specific cellular responses.** (Top) Gpr124 and Reck cooperatively alter the cell's perception of its Wnt microenvironment by selectively potentiating Wnt7 signals (cyan-tinted dots). (Bottom) Reck decodes Wnt ligands by establishing monospecific contacts with the highly divergent Wnt7 linker domain. Gpr124 links Reck-bound Wnt7 to Dishevelled. Dishevelled polymers by interacting simultaneously with Gpr124 and Fz assemble Wnt7-enriched signalosomes that trigger signaling through Fz receptors and Lrp5/6 co-receptors.

The list of author affiliations is available in the full article online.

\*These authors contributed equally to this work.

†Corresponding author. Email: [benoit.vanhollebeke@ulb.ac.be](mailto:benoit.vanhollebeke@ulb.ac.be)  
Cite this article as M. Eubelen et al., *Science* 361, eaat1178 (2018). DOI: 10.1126/science.aat1178

## RESEARCH ARTICLE

## DEVELOPMENTAL SIGNALING

# A molecular mechanism for Wnt ligand-specific signaling

Marie Eubelen<sup>1\*</sup>, Naguissa Bostaille<sup>1\*</sup>, Pauline Cabochette<sup>1</sup>, Anne Gauquier<sup>1</sup>, Patricia Tebabi<sup>1</sup>, Andra C. Dumitru<sup>2</sup>, Melanie Koehler<sup>2</sup>, Philipp Gut<sup>1†</sup>, David Alsteens<sup>2</sup>, Didier Y. R. Stainier<sup>3</sup>, Abel Garcia-Pino<sup>4,5</sup>, Benoit Vanhollebeke<sup>1,5,6‡</sup>

Wnt signaling is key to many developmental, physiological, and disease processes in which cells seem able to discriminate between multiple Wnt ligands. This selective Wnt recognition or “decoding” capacity has remained enigmatic because Wnt/Frizzled interactions are largely incompatible with monospecific recognition. Gpr124 and Reck enable brain endothelial cells to selectively respond to Wnt7. We show that Reck binds with low micromolar affinity to the intrinsically disordered linker region of Wnt7. Availability of Reck-bound Wnt7 for Frizzled signaling relies on the interaction between Gpr124 and Dishevelled. Through polymerization, Dishevelled recruits Gpr124 and the associated Reck-bound Wnt7 into dynamic Wnt/Frizzled/Lrp5/6 signalosomes, resulting in increased local concentrations of Wnt7 available for Frizzled signaling. This work provides mechanistic insights into the Wnt decoding capacities of vertebrate cells and unravels structural determinants of the functional diversification of Wnt family members.

**W**nts constitute a large family of highly conserved and secreted proteins that mediate intercellular communication during animal development and in adult tissue homeostasis (1, 2). The 10 members of the Frizzled (Fz) family are seven-pass transmembrane proteins that serve as receptors for Wnts (3–5). Greatly contributing to the complexity of Wnt signaling, the Wnt/Fz binding relationships are promiscuous, with multiple Wnts competing for binding to individual Fzs and vice versa (3–12). The Wnt/Fz contacts are mediated by conserved residues or common chemical modifications (13). These observations raise the question of how cells can respond to specific Wnt ligands when exposed to the overlapping expression patterns of multiple Wnt ligands that sometimes have opposing biological functions.

A pertinent example is the exclusive control of mammalian forebrain and ventral spinal cord angiogenesis by Wnt7a and Wnt7b (14–16). Spe-

cifically, in order to respond to neural progenitor-derived Wnt7 by activating Wnt/ $\beta$ -catenin signaling, endothelial cells must express Gpr124, an orphan member of the adhesion class of G protein-coupled receptors (GPCRs) (17–22) as well as the glycosylphosphatidylinositol (GPI)-anchored glycoprotein Reck (22, 24). Gpr124 and Reck physically interact to synergistically stimulate Wnt7-specific responses (22, 24), but it is unknown how Wnt7 signals are specifically recognized and transduced.

## Reck is a Frizzled-independent Wnt7-specific receptor

We first sought to determine the Wnt7 recognition mechanism, a question inherently complicated by the ubiquitous expression of Fz receptors and their Lrp5/6 co-receptors in vertebrate cells. We therefore generated a set of mutant human embryonic kidney (HEK)-293 cell lines by targeting (i) all 10 FZ genes ( $FZ_{1-10}^{-/-}$ ), (ii)  $LRP5$  and/or  $LRP6$ , or (iii)  $GPR124$  and  $RECK$ , through multiplexed CRISPR/Cas9 mutagenesis (Materials and methods, Fig. 1A, and figs. S1 to S3). Ectopically expressed V5-tagged Wnt7a (Wnt7a-V5) could be immunodetected at the plasma membrane of wild-type (WT),  $FZ_{1-10}^{-/-}$ , and  $LRP5^{-/-}$ ;  $LRP6^{-/-}$  cells but not  $GPR124^{-/-}$ ;  $RECK^{-/-}$  cells (Fig. 1B). Ectopic restoration of Reck, alone or in combination with Gpr124, was sufficient to restore Wnt7a-V5 membrane labeling of  $GPR124^{-/-}$ ;  $RECK^{-/-}$  cells (Fig. 1C). Fz5 also bound Wnt7a-V5, reflecting the competence of this receptor to mediate baseline Wnt7 signaling (11, 24). Control Wnt3a-V5 did not label WT or mutant cells (Fig. 1B and fig. S4).

Proximity ligation assays (PLAs) allow localized detection of protein interactions at the single-

molecule level through DNA rolling-circle signal amplification. PLAs in  $GPR124^{-/-}$ ;  $RECK^{-/-}$  cells confirmed the interaction between Wnt7a-V5 and hemagglutinin (HA)-Reck or HA-Fz5 at the plasma membrane. The fraction of PLA-positive cells was identical ( $28.8 \pm 7.5\%$  and  $28.3 \pm 7.0\%$ , respectively), but the intensity of the PLA signals generated with HA-Reck was twofold higher than with HA-Fz5 (Fig. 1D) (quantification protocol is provided in Materials and methods). By contrast, HA-Gpr124 did not generate PLA signals. We tested all the Wnt family members and found that PLA interaction signals with HA-Reck were restricted to Wnt7a-V5 and Wnt7b-V5 (Fig. 1E), reflecting the specificity of the Gpr124/Reck complex for Wnt7 signaling (20–22, 24).

We next determined whether Reck recruits Wnt7a in the absence of functional Fz receptors. We first performed PLAs between V5-tagged Wnt7a or Wnt3a secreted from WT cells and HA-Reck exposed at the surface of neighboring green fluorescent protein (GFP)-labeled  $FZ_{1-10}^{-/-}$  cells. PLA signals were detected at the plasma membrane of  $57.7 \pm 17.4\%$  of the GFP<sup>+</sup> cells in Wnt7a<sup>+</sup> cocultures but were undetectable in Wnt3a<sup>+</sup> cocultures (Fig. 1F). Accordingly, the coculture of  $FZ_{1-10}^{-/-}$  cells expressing Reck but not  $Gpr124^{\Delta ICD}$  could drastically reduce Wnt7a-Fz5 signaling in neighboring Super Top Flash (STF) reporter cells (Fig. 1G). By contrast, none of the other tested Wnts, including Wnt3a (Fig. 1G), could be trapped by Reck-expressing cells (fig. S5). In this ligand capture assay, Gpr124 was lacking most of its C-terminal intracellular domain (ICD) domain (residues 81 to 337 of the ICD) to restrict the analysis to the extracellular parts of the Gpr124/Reck complex.

## Defining the Reck-Wnt7 monospecific recognition mechanism

Next, we mapped the domains of Reck required for Wnt7a binding. We generated a collection of HA-tagged single-domain deletion variants of Reck and, after determining that all variants reached the plasma membrane (fig. S6), quantified Wnt7 PLA interaction signals. This analysis revealed that the N-terminal cysteine-knot domain (CK), and more particularly CK4 and CK5, were required for binding (Fig. 2A). Consistent with these results, Reck <sup>$\Delta$ CK4</sup> was also inactive in competition assays (Fig. 1G).

Reck function in Wnt signaling is known to rely on its capacity to form a complex with Gpr124 through its CK domain (22, 24), which we show to be required for Wnt7a binding. Although Gpr124 alone did not bind Wnt7 (Fig. 1, C, D, and G), we detected a fourfold increase in Wnt7a-V5/HA-Reck PLA signals upon coexpression of untagged Gpr124 (Fig. 2B). These results suggest that Reck is a Fz-independent Wnt receptor, whose specific and exclusive binding to Wnt7 is reinforced by the interaction of Gpr124 with its CK domain (Fig. 2C) (24).

In order to better characterize the Reck/Wnt7 interaction, we modeled the three-dimensional (3D) structure of Wnt7a based on *Xenopus* Wnt8a crystallographic analysis (Fig. 2D) (13). Wnt

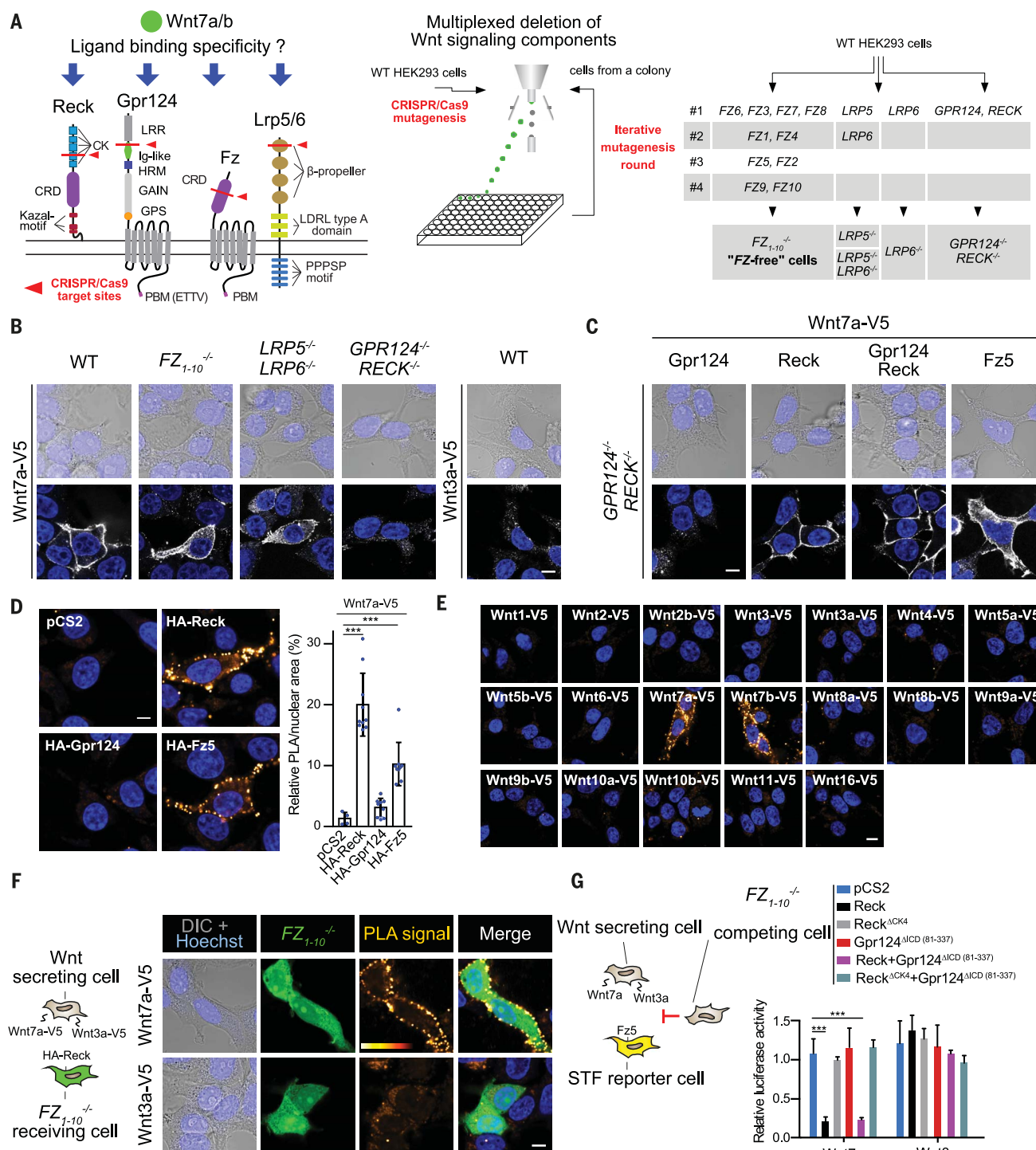
<sup>1</sup>Laboratory of Neurovascular Signaling, Department of Molecular Biology, ULB Neuroscience Institute, Université libre de Bruxelles (ULB), Gosselies B-6041, Belgium.

<sup>2</sup>NanoBiophysics Laboratory, Louvain Institute of Biomolecular Science and Technology, Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium. <sup>3</sup>Department of Developmental Genetics, Max Planck Institute for Heart and Lung Research, 61231 Bad Nauheim, Germany. <sup>4</sup>Laboratory of Cellular and Molecular Microbiology, Department of Molecular Biology, ULB, Gosselies B-6041, Belgium. <sup>5</sup>Walloon Excellence in Life Sciences and Biotechnology (WELBIO), Belgium. <sup>6</sup>Center for Microscopy and Molecular Imaging (CMMI), ULB, Gosselies B-6041, Belgium.

\*These authors contributed equally to this work. †Present address: Nestlé Institute of Health Sciences, EPFL Innovation Park, Lausanne, Switzerland.

‡Corresponding author. Email: benoit.vanhollebeke@ulb.ac.be

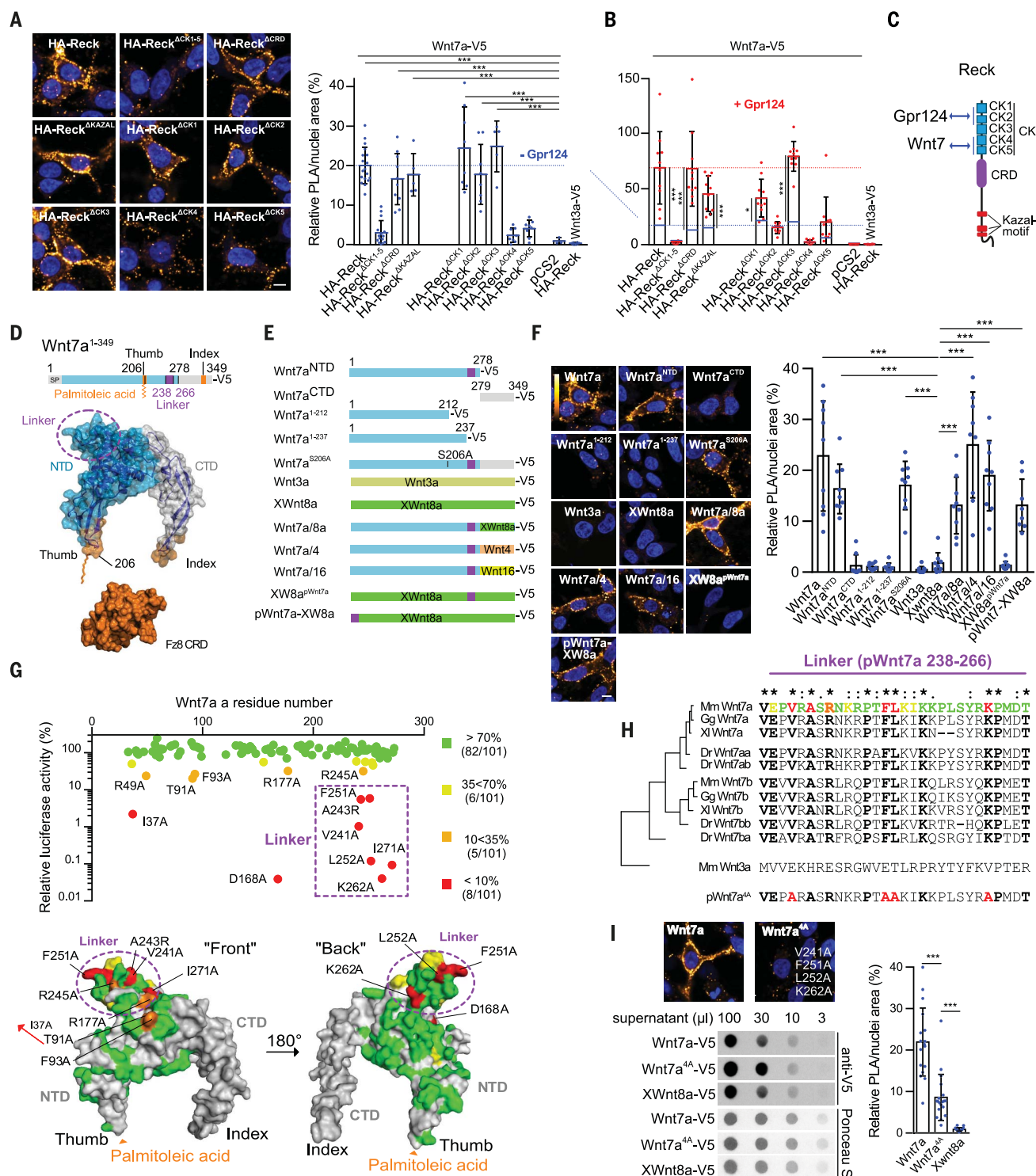




**Fig. 1. Reck is a Frizzled-independent Wnt7-specific receptor.**

(A) Schematics of the Wnt7 receptor complex components and strategy for their genetic inactivation. Cells transfected with a bicistronic construct encoding the sgRNA and SpCas9-2A-GFP were iteratively cloned by means of fluorescence-activated cell sorting (FACS) (49). (B) Anti-V5 immunostaining of transiently expressed Wnt7a-V5 or Wnt3a-V5 in cells of the indicated genotype. (C) Same as (B) for cells transiently coexpressing Wnt7a-V5 and the indicated receptors and co-receptors. (D) PLA using antibody to V5 and antibody to HA directed against Wnt7a-V5 and the indicated HA-tagged receptors and co-receptors expressed in *GPR124*<sup>-/-</sup>; *RECK*<sup>-/-</sup> cells. Signal quantification

is on the right (Materials and methods). (E) Anti-V5/HA PLAs between the indicated V5-tagged Wnt ligand and HA-Reck in *GPR124*<sup>-/-</sup>; *RECK*<sup>-/-</sup> cells. (F) Anti-V5/HA PLAs in (1:1) cocultures of Wnt7a-V5- or Wnt3a-V5-secreting WT cells and *FZ1-10*<sup>-/-</sup> HEK293T cells coexpressing HA-Reck and cytosolic GFP as a transfection marker. (G) Ligand capture assay in triple (1:1:1) cocultures. Luciferase activities of HEK293-STF reporter cells transfected with Fz5 (yellow cell) and stimulated in a paracrine manner by cocultured Wnt7a- or Wnt3a-secreting cells in the presence of *FZ1-10*<sup>-/-</sup> HEK293T competing cells transfected with various receptors and co-receptors as indicated. Scale bars, 10  $\mu$ m. \*\*\**P* < 0.001; data represent mean  $\pm$  SD.



**Fig. 2. Wnt7 recognition involves its intrinsically disordered linker region.** (A) Anti-V5/HA PLAs between Wnt7a-V5 and HA-Reck or its variants coexpressed in *GPR124*<sup>-/-</sup>/*RECK*<sup>-/-</sup> cells in the absence of Gpr124. Quantification is shown on the right. (B) Quantification as in (A) in the presence of Gpr124. (C) Reck and its Gpr124 and Wnt7 binding partners. (D) Structural modeling of Wnt7a based on the crystal coordinates of XWnt8a. The NTD and CTD are pseudocolored in cyan and gray, respectively. The NTD-embedded linker region is circled in purple. (E) Schematic representation of the V5-tagged Wnt variants examined in the anti-V5/HA PLAs shown and quantified in (F) in *GPR124*<sup>-/-</sup>/*RECK*<sup>-/-</sup> cells coexpressing HA-Reck. (G) Gpr124/Reck-dependent luciferase activities of 101 different single-residue variants of Wnt7a normalized to WT Wnt7a in HEK293-STF cells. (H) Alignment of the Wnt7 linker domain across the vertebrate clade. Mm, *Mus musculus*; Gg, *Gallus gallus*; Xi, *Xenopus laevis*; Dr, *Danio rerio*. The residues of Mm Wnt7a are color-coded according to their activity class determined in (G). (I) Anti-V5/HA PLAs between Wnt7a-V5 or Wnt7a<sup>ΔA</sup>-V5 and HA-Reck. Ligand secretion was evaluated by means of semiquantitative anti-V5 dot blot analysis of serially diluted cell supernatant. Ponceau S staining was used as the loading control. Scale bars, 10 μm. \**P* < 0.05, \*\*\**P* < 0.001; data represent mean ± SD.

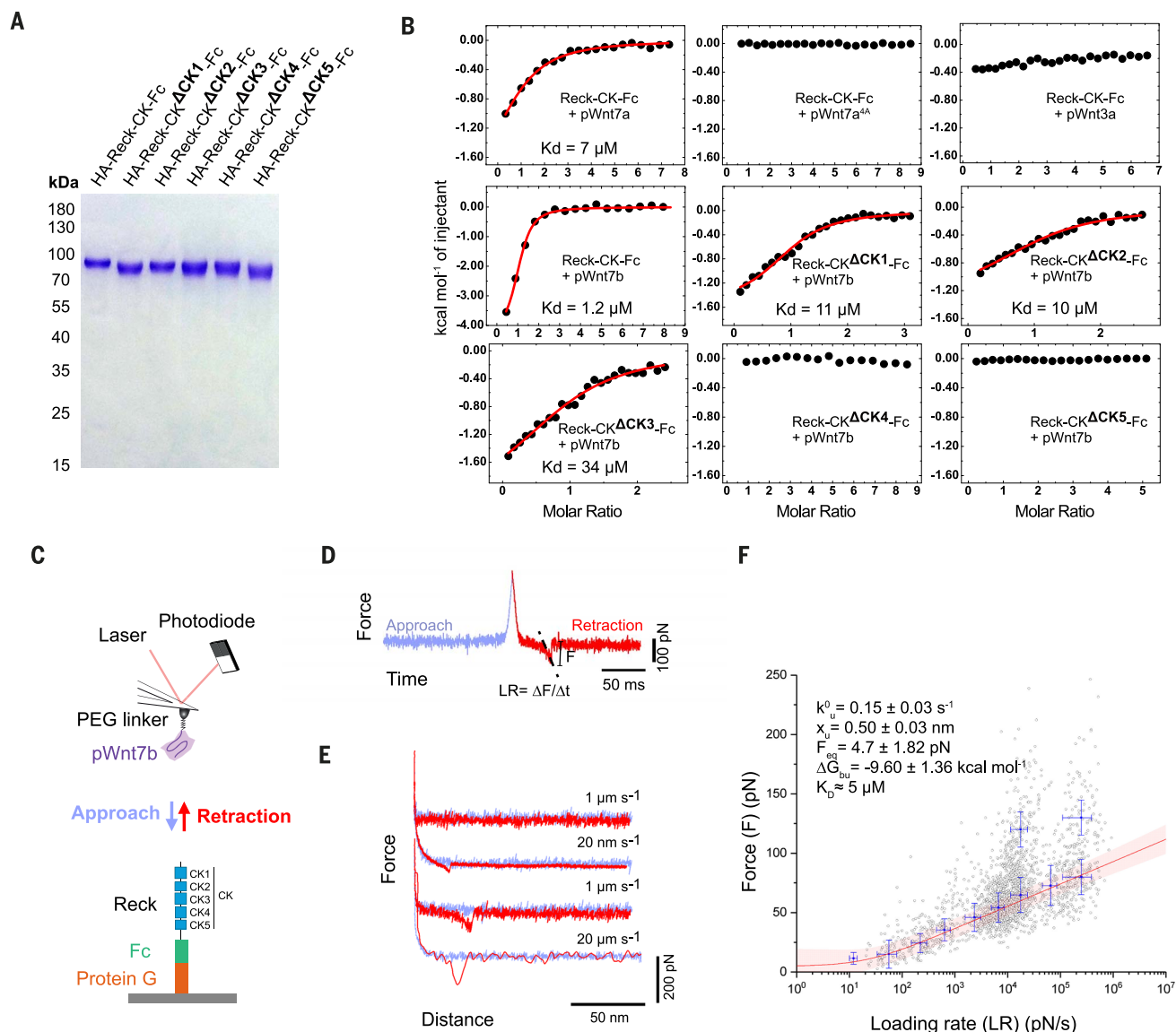
cells coexpressing HA-Reck. (G) Gpr124/Reck-dependent luciferase activities of 101 different single-residue variants of Wnt7a normalized to WT Wnt7a in HEK293-STF cells. (H) Alignment of the Wnt7 linker domain across the vertebrate clade. Mm, *Mus musculus*; Gg, *Gallus gallus*; Xi, *Xenopus laevis*; Dr, *Danio rerio*. The residues of Mm Wnt7a are color-coded according to their activity class determined in (G). (I) Anti-V5/HA PLAs between Wnt7a-V5 or Wnt7a<sup>ΔA</sup>-V5 and HA-Reck. Ligand secretion was evaluated by means of semiquantitative anti-V5 dot blot analysis of serially diluted cell supernatant. Ponceau S staining was used as the loading control. Scale bars, 10 μm. \**P* < 0.05, \*\*\**P* < 0.001; data represent mean ± SD.



ligands adopt a two-domain structure reminiscent of a human hand pinching the globular Fz cysteine-rich domain (CRD) (Fig. 2D, orange) via the palmitoylated “thumb” of their N-terminal domain (NTD) (Fig. 2D, cyan) and hydrophobic residues of their “index” C-terminal domain (CTD) (Fig. 2D, gray). The structures involved in Fz binding are pseudocolored in orange in Fig. 2D. The two domains are connected through a flexible linker region of the NTD (Fig. 2D, purple).

We next generated a collection of Wnt7 variants deleted for specific domains or residues or that carried domains from other Wnt ligands (Fig. 2E). All analyzed Wnt7 variants could be detected in the cell supernatant (fig. S7). These ligands were each applied to PLAs, and the resulting data showed that Reck binding occurs through the Wnt7a NTD. Wnt7a<sup>NTD</sup> indeed bound Reck, whereas Wnt7a<sup>CTD</sup> did not (Fig. 2F). Chimeric ligands made of Wnt7a<sup>NTD</sup> fused to the

CTDs of *Xenopus* Wnt8a (XWnt8a) or mouse Wnt4 or Wnt16 were also positive. The palmitoleic acid required for Fz binding was dispensable for the interaction with Reck, as revealed through testing a Wnt7a<sup>S206A</sup> palmitoylation mutant. Altogether, these binding assays reveal that Reck discriminates between Wnt ligands by recognizing a motif embedded in the Wnt7a NTD at sites distinct from those engaged by Fz.



**Fig. 3. The Wnt7 linker peptide binds the Reck CK domain with low micromolar affinity.** (A) Coomassie Blue staining of the recombinant Reck-CK-Fc fusion proteins used in biophysical analyses. (B) Interaction between Reck-CK-Fc and synthetic pWnt7 peptides by use of isothermal titration calorimetry (ITC). (Top) ITC of pWnt7a, pWnt7a<sup>4A</sup>, and pWnt3a into Reck-CK-Fc. (Middle and bottom) ITC of Reck-CK-Fc and its CK motif variants to pWnt7b. (C) Probing Reck-CK-Fc/pWnt7b interaction by use of SMFS. Principle of force-distance curve-based atomic force microscopy (FD-based AFM) is shown. An AFM tip functionalized with a polyethylene glycol (PEG) spacer fused to the pWnt7b peptide is approached and retracted from a surface coated with Reck-CK-Fc. (D) Force-time curve from which the loading

rate (LR) can be extracted from the slope of the curve just before bond rupture ( $LR = \Delta F/\Delta t$ ). (E) Force-distance curves recorded at various retraction speeds (from 20 nm s<sup>-1</sup> to 20 μm s<sup>-1</sup>) and showing no adhesion event (first curve) or specific adhesion events (other curves). (F) Loading rate-dependent interaction forces of single Wnt7b peptide-receptor bonds quantitate the ligand-binding energy landscape of Reck. (Inset) Fitting the data by using the Fiddle-Noy-de Yoreo model (thin red line) (54) provides average equilibrium force ( $F_{eq}$ ), barrier location of the unbound state ( $x_u$ ), equilibrium free energy between the bound and unbound state ( $\Delta G_{bu}$ ), intrinsic unbinding rate ( $k_u^0$ ), and  $K_D$ , with errors representing the S.E.. Each circle represents one measurement. Darker shaded areas represent 95% prediction interval.

Smaller NTD variants lacking the linker region, Wnt7a<sup>1-212</sup> and Wnt7<sup>1-237</sup>, did not bind Reck. This mutational analysis and the spatial segregation of the Wnt7a linker region from the Fz binding sites suggest that Reck decodes Wnt7a, at least in part, through this linker. The linker, which is highly divergent among the different Wnt ligands, exhibits strong evolutionary conservation among the vertebrate Wnt7 orthologs (Fig. 2H and fig. S8). This region is predicted to be intrinsically disordered (fig. S9), a feature often found in molecular recognition elements providing the necessary structural plasticity to accommodate multiple partners, post-translational modifications, or moonlighting functions. In many cases, intrinsically disordered regions also provide interactions with high specificity and moderate-to-high affinity (25).

Although in situ substitution of the XWnt8a linker by the Wnt7 linker (XW8a<sup>pWnt7a</sup>) did not yield detectable PLA signals, presenting the Wnt7 linker at the free N terminus of XWnt8a (pWnt7a-XW8a) was sufficient to confer Reck binding activity to XWnt8a (Fig. 2, E and F). We hypothesize that conformational alteration accounts for the lack of binding of XW8a<sup>pWnt7a</sup>. In support of this hypothesis, reciprocal exchange of Wnt linkers between various Wnts abrogates their activity (fig. S10).

To precisely map the Reck interaction site, we analyzed Gpr124/Reck-dependent STF signaling of 101 single-residue variants of Wnt7a (Fig. 2G). The mutated residues correspond to surface-exposed NTD residues conserved between Wnt7a and Wnt7b but not found in XWnt8a or other Wnt ligands. Residues were mutated to alanines, except for endogenous alanine residues, which were changed to arginine. Although ~80% of the variants were as active as WT Wnt7a (>70% relative activity) (Fig. 2G, green), eight Wnt7a variants (Fig. 2G, red) reduced Gpr124/Reck-dependent signaling to <10%. All but one (I37) critical residues clustered on the “top” or “back” of the predicted Wnt7a structure, with six mapping to the linker domain. All essential residues are strictly conserved among Wnt7 orthologs from fish to mammals and absent in other Wnts, including Wnt3a (Fig. 2H and fig. S8).

The linker domain of Wnt3a has been shown to be essential for Wnt3a activity through Lrp6 binding (26). By analogy, the inactivity of the Wnt7 linker variants might therefore result from defective binding to Lrp5/6, Reck, or both. In line with a function in Reck binding, Wnt7a<sup>4A</sup>, a four-residue variant of Wnt7a (V241A/F251A/L252A/K262A) within the linker region (Fig. 2H), showed reduced Reck PLA signals as compared with that of WT Wnt7a. This lower activity occurred despite slightly improved secretion rates (Fig. 2I).

### Biophysical characterization of the interaction between the Wnt7 linker domain and Reck

To investigate Reck-Wnt7 binding in a cell-free system, the CK domain of Reck (and variants thereof) were fused to the Fc domain of human immunoglobulin G1 (IgG1). Fusion proteins were

purified from HEK293T cells supernatants (Fig. 3A) and then titrated with synthetic pWnt7a and pWnt7b linker peptides by means of isothermal titration calorimetry (ITC). pWnt7a and pWnt7b bound Reck with affinity values of 7 and 1.2  $\mu$ M, respectively (Fig. 3B and fig. S11). As controls, the synthetic peptides corresponding to Wnt7a<sup>4A</sup> (pWnt7a<sup>4A</sup>) as well as equivalent linker peptides of Wnt3a (pWnt3a) showed no binding to Reck-CK-Fc. pWnt7b binding required Reck CK4 and CK5 but not CK1, CK2, or CK3, mirroring the PLA results in cultured cells (Fig. 2A). To corroborate the results provided by the ITC analysis, we used single-molecule force spectroscopy (SMFS) to measure binding affinities at the single-molecule level (Fig. 3C). Binding of pWnt7b to Reck-CK-Fc was detectable with a measured dissociation constant ( $K_d$ ) of 5  $\mu$ M (Fig. 3, D to F). Despite the fundamental differences between the two techniques, ITC and SMFS thus provided a close match between measured binding affinity values.

Altogether, although not excluding an additional role of the Wnt7 linker in Lrp5/6 binding (26), these data demonstrate that Wnt7 is recognized by Reck at least in part through its “signature” linker motif. The moderate micromolar affinity values measured further suggest that after recognition of the linker, Reck establishes more extensive contacts with Wnt7, in a process that can be potentiated by Gpr124 (Fig. 2B).

### Gpr124 function in Wnt7 signaling does not depend on its GPCR structure

Reck, by virtue of its GPI-anchoring mode, has limited potential to relay Wnt7 signals within the cell. Signal transduction therefore likely relies on other components of the receptor complex—Gpr124 and/or Fz/Lrp5/6 (Fig. 4A).

To uncover the signal transduction mechanism, we first evaluated the functional relationship between Gpr124 and Fz/Lrp5/6 complexes in cultured cells. Using the Fz and Lrp mutant cells (Fig. 1A), we found that the function of Gpr124/Reck strictly relies on Fz and Lrp5/6 (fig. S12). We further established that their respective CRD and Dkk-1-sensitive Wnt ligand-binding domains are essential, implying that Wnt7 binds and activates Fz/Lrp5/6 in a classical manner. If Reck mediates Wnt7 binding and Fz/Lrp5/6 trigger signaling, what underlies the essential function of Gpr124?

In the absence of Gpr124, Reck acts cell-autonomously as a potent inhibitor of Wnt7/Fz5 signaling, in a CK4-dependent manner (Fig. 4B). The inhibitory function of standalone Reck is particularly remarkable in light of the pool of membrane-associated Wnt7 in this Frizzled-positive setting (Fig. 4B). This observation implies that in the absence of Gpr124, Reck scavenges Wnt7 away from Fz5/Lrp5/6 complexes. The presence of Gpr124 switches Wnt7 signaling output from near complete inhibition to potent activation.

To establish how Gpr124 mediates this “on-off” Wnt7 signaling switch, we turned to the zebrafish model. In many Wnt-controlled processes, includ-

ing the Wnt7/Gpr124/Reck-mediated cerebrovascular functions, Wnt input levels are only marginally above the minimal threshold values required for signaling (20). It is therefore important to investigate the signal transduction pathway in vivo, in response to physiological Wnt7 inputs. The development of the zebrafish brain vasculature requires Reck/Gpr124 signaling in a process of angiogenic sprouting that can readily be quantified. It therefore constitutes an ideal setting in which to perform structure-function analysis in vivo (22, 27). Using mRNA injections into one-cell stage *gpr124*<sup>-/-</sup> embryos (Fig. 4C), we evaluated the activity of three Gpr124 variants lacking the N-terminal extracellular part (Gpr124<sup>ΔECD</sup>), the seven-span moiety (Gpr124<sup>ΔTM2-7</sup>), or the C-terminal cytoplasmic extension (Gpr124<sup>ΔICD</sup>, lacking residues 29 to 337 of the ICD) (Fig. 4D). Although ectopic expression of Gpr124<sup>ΔECD</sup> or Gpr124<sup>ΔICD</sup> did not restore brain angiogenesis in *gpr124* mutants or morphants, Gpr124<sup>ΔTM2-7</sup> was sufficient to trigger brain angiogenesis in vivo (Fig. 4, E and F, and fig. S13) and Wnt/ $\beta$ -catenin activity in vitro (fig. S14).

This retained competence of Gpr124<sup>ΔTM2-7</sup> was unexpected: Gpr124 is a GPCR, a receptor superfamily classically relaying extracellular stimuli through ligand-induced conformational remodeling of their seven transmembrane spans, which are absent in the engineered Gpr124<sup>ΔTM2-7</sup>. These data raise the possibility that Gpr124 does not act as a “classical” GPCR when promoting Wnt7 signaling.

To test this hypothesis, we developed a bi-molecular complementation assay in which the GPR124 extracellular domain (ECD) and ICD are linked by a surrogate anti-GFP VhhGFP4 nanobody-GFP connector (Fig. 4G). Nanobodies are single-domain antibody fragments that have been used to rout or misroute intracellular proteins (28). We repurposed them here as conditional tethers for signal transduction analysis. The highly flexible GFP-VhhGFP4 connector acts a buffering module, ensuring that conformational information cannot be exchanged between tethered partners. On the basis of this idea, we designed Gpr124<sup>ΔICD</sup>-GFP and Gpr125<sup>ΔICD</sup>-GFP fusions to which VhhGFP4 fusions will be recruited (Fig. 4G and fig. S15). The *gpr124* vascular phenotypes were partially suppressed by co-injecting mRNAs encoding Gpr124<sup>ΔICD</sup>-GFP and VhhGFP4-ICD (residues 29 to 337) (Fig. 4H). Moreover, a chimeric Gpr124 ECD linked to cytoplasmic GFP via the transmembrane span of the unrelated CD27 receptor was similarly active with VhhGFP4-ICD (Fig. 4, G and H). We used Gpr125, a closely related aGPCR devoid of angiogenic activity, as well as VhhGFP4-red fluorescent protein (RFP) as negative controls (20–22).

These results confirm that Gpr124 function in Wnt7 signaling does not require signal transduction across the membrane through conformational remodeling. Instead, Gpr124 seemingly acts in this module as a signaling-deficient transmembrane protein whose activity relies on its Reck-binding ECD (22, 24) and its conformationally



uncoupled ICD. The function of this latter domain remains to be defined (Fig. 4I).

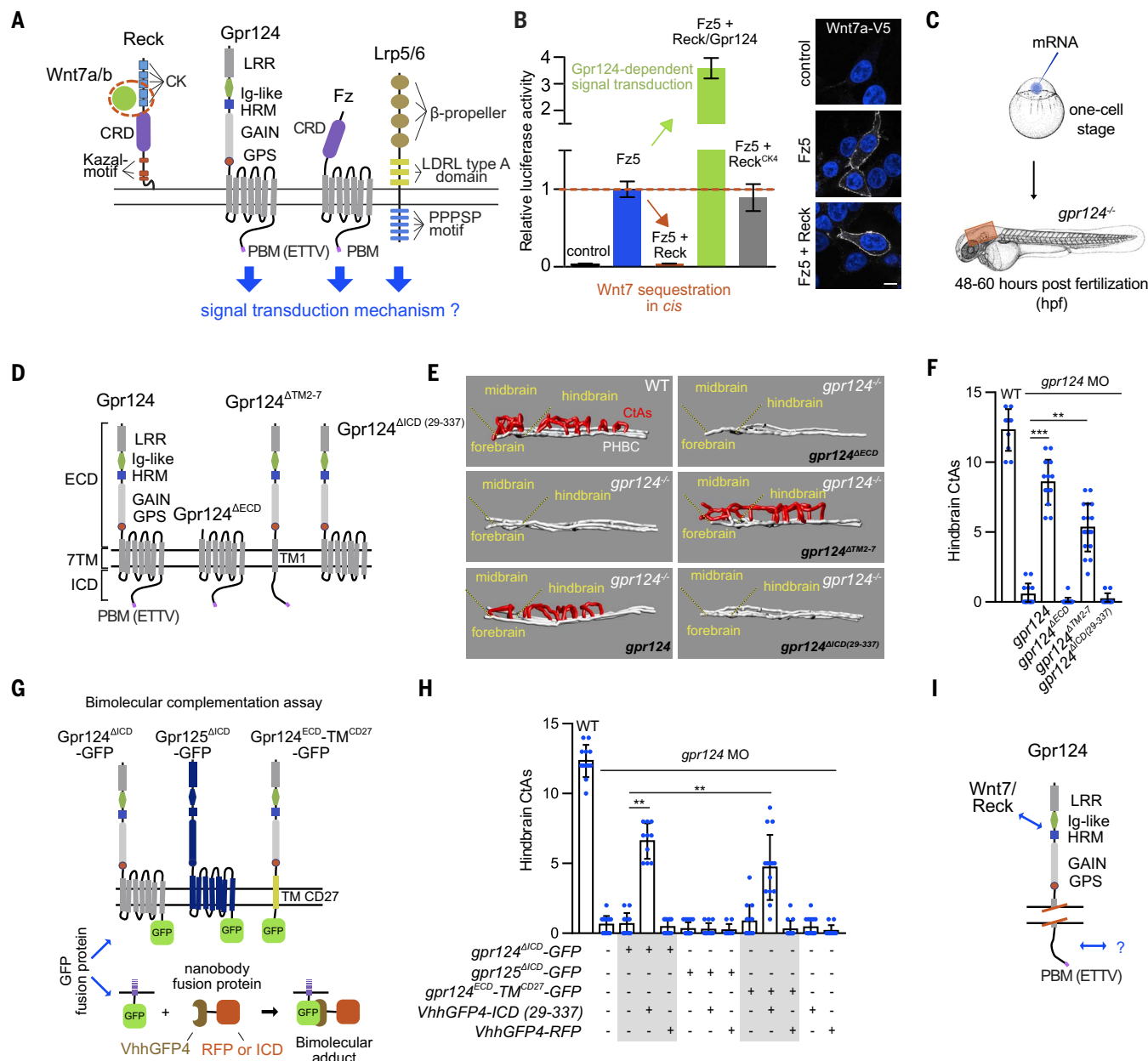
### Gpr124 activity requires Dishevelled binding

We hypothesized that the Gpr124 ICD might operate through Dishevelled (Dvl), the necessary

effector of Wnt signaling that interacts with Fz. This “Dvl hypothesis” is rooted in the findings that Gpr125 physically interacts with Dvl via its C-terminal ICD domain (29) and that Gpr124/125 hybrids in which the ICD of Gpr124 is replaced with the ICD of Gpr125 are able to promote brain angiogenesis (Fig. 5A) (22). Gpr124/Fz2 hy-

brids (Gpr124<sup>ICDFz2</sup>) harboring the Fz2 ICD, which is known to bind Dvl, were similarly active. By contrast, full-length Fz2 was not. The activity of Gpr124<sup>ICDFz2</sup> was dependent on its KTxxxW and ETTV Dvl binding motifs (Fig. 5A).

Moreover, in a noncanonical Wnt signaling context, overexpression of Gpr125 impairs



**Fig. 4. Gpr124 function in Wnt signaling does not depend on its GPCR structure.** (A) Potential Wnt7-specific signal transduction mechanisms. (B) Luciferase activity (left) and anti-V5 immunostaining (right) of HEK293-STF cells cotransfected with Fz5, Wnt7a, and the indicated receptors and co-receptors. (C) Experimental setup for in vivo brain angiogenesis assays in zebrafish embryos after RNA injections at the one-cell stage. The red area is imaged to analyze the hindbrain central arteries (CtAs). (D) Gpr124 and its domain variants. (E) Representative 3D wire diagrams of the cerebrovasculature of WT and *gpr124* mutant embryos at 60 hours post fertilization, injected with 100 pg of the indicated RNA. Red vessels represent the Gpr124/Reck-dependent intracerebral CtAs that sprout from the gray perineural

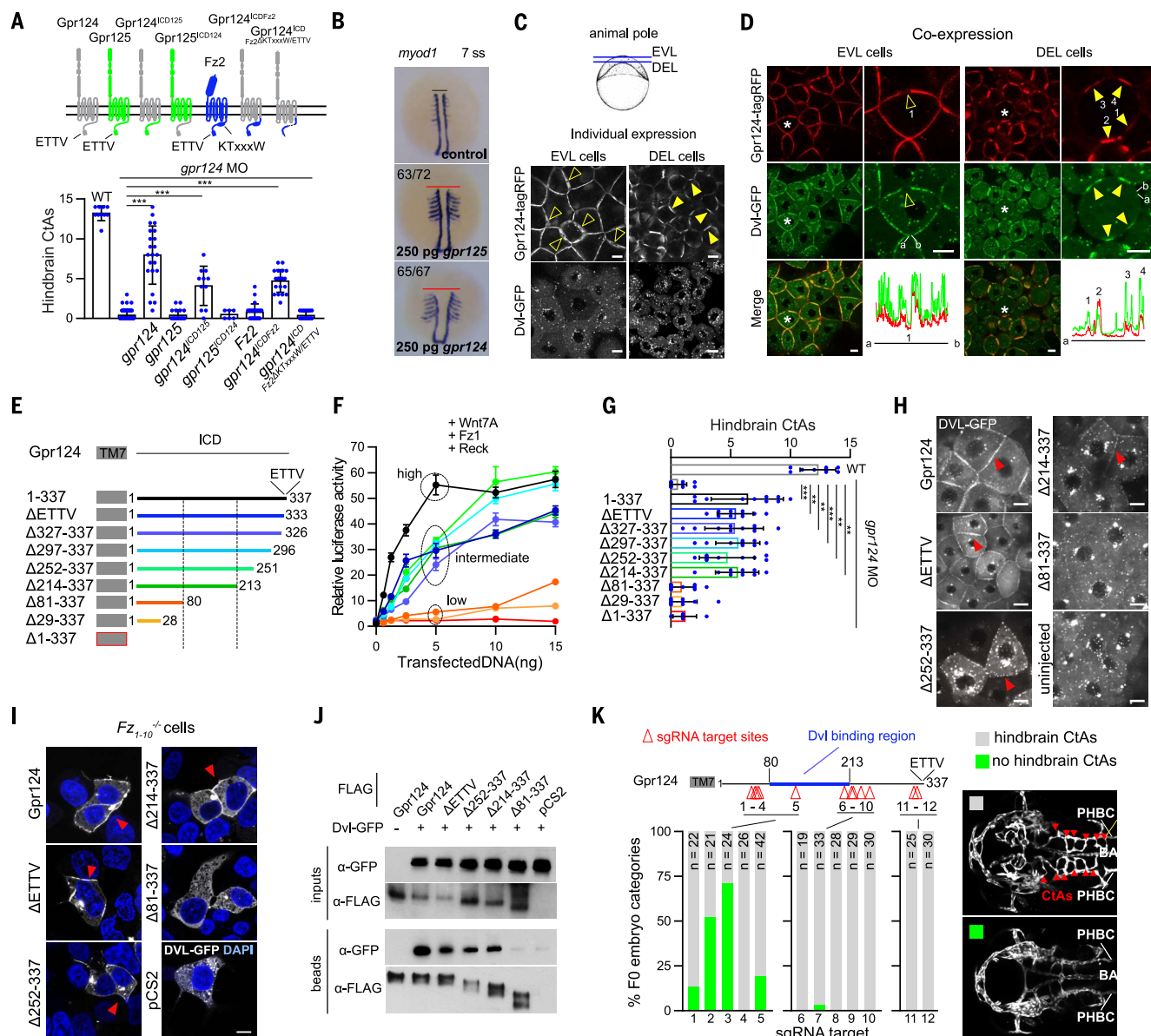
primordial hindbrain channels (PHBCs); the confocal stacks used to generate the wire diagrams are available in fig. S13. (F) Hindbrain CtAs of *gpr124* morphant embryos at 60 hours post fertilization, injected at the one-cell stage with 100 pg of the indicated mRNA. (G) Bimolecular GPCR complementation assay strategy. Membrane receptors (Gpr124, Gpr125, and Gpr124<sup>ECD-TM<sup>CD27</sup></sup>) inactivated by ICD/GFP substitutions are coexpressed with VhhGFP4 fusions to reconstitute a bimolecular adduct through an artificial GFP-VhhGFP4 linker. (H) Hindbrain CtAs of *gpr124* morphant embryos at 60 hours post fertilization, injected at the one-cell stage with 100 pg of the indicated mRNAs. (I) Gpr124 and its essential protein interactions. Scale bar, 10 μm. \*\**P* < 0.01, \*\*\**P* < 0.001; data represent mean ± SD.

convergence and extension movements during zebrafish gastrulation, resulting in a mediolateral broadening of the anterior-posterior embryonic axis as well as synophthalmia or cyclopia (29). These phenotypes were linked to the capacity

of Gpr125 to modulate Dvl distribution through its ICD. Gpr124 strictly mimicked Gpr125 in these settings (Fig. 5B and fig. S16).

We therefore tested whether Gpr124 affects the intracellular distribution of Dvl. We used

two distinct cell populations of the zebrafish blastula to address this question: the superficial enveloping layer (EVL) cells, which maintain continuous intercellular contacts, and the deep layer (DEL) cells, which establish more discrete



**Fig. 5. Gpr124 interacts with Dvl through its ICD.** (A) Hindbrain CtAs of *gpr124* morphant embryos at 60 hours post fertilization, injected at the one-cell stage with 100 pg mRNA encoding the illustrated receptors. (B) WISH of *myod1* expression at the seven-somite stage embryos, control, or injected with 250 pg of *gpr125* or *gpr124* mRNA. The red lines indicate the mediolateral broadening of the anterior-posterior axis. (C) Intracellular distribution of Gpr124-tagRFP and Dvl-GFP expressed individually in zebrafish blastulae. (D) Intracellular distribution of Gpr124-tagRFP and Dvl-GFP coexpressed in zebrafish blastulae. The cells annotated with an asterisk are magnified at right, and the pixel intensity of the green and red channels along a virtual clockwise path following the cell cortex from a to b is plotted below. In (C) and (D), open and solid arrowheads point respectively to Gpr124-negative and Gpr124-positive membrane microdomains in EVL and DEL cells. (E) Gpr124 and its ICD variants. (F) Luciferase dose-response

values of HEK293-STF cells cotransfected with Wnt7a, Fz1, Reck, and increasing amounts of Gpr124 constructs or its ICD variants as illustrated in (E). (G) Hindbrain CtAs of *gpr124* morphant embryos at 60 hours post fertilization, injected at the one-cell stage with 100 pg of the indicated mRNA. (H) Intracellular distribution of Dvl-GFP coexpressed with Gpr124 and its ICD variants in zebrafish blastula EVL cells. (I) same as (H) in *FZ1-10*<sup>-/-</sup> cells. In (H) and (I), red arrowheads point to Dvl-GFP signal at the plasma membrane. (J) Anti-FLAG coimmunoprecipitation assays in total lysates of cells coexpressing Dvl-GFP and N-terminal FLAG-tagged versions of Gpr124 or its ICD variants. (K) Mosaic gene disruption in somatic zebrafish embryos by means of injection at the one-cell stage of 150 pg of *zCas9* mRNA and 50 pg of the illustrated sgRNAs. Resulting hindbrain vasculature scored at 48 hours post fertilization. BA, Basilar artery. Scale bars, 10  $\mu$ m. \*\**P* < 0.01, \*\*\**P* < 0.001; data represent mean  $\pm$  SD.



intercellular junctions (Fig. 5C). Gpr124-tagRFP distribution reflects this differing junctional organization, with near-uniform membrane signals in EVL and discontinuous signals in DEL cells (Fig. 5C). When expressed individually, *Xenopus* Dvl-GFP mainly formed cytoplasmic punctae in both cell populations, as reported previously (29). Coexpressing Gpr124-tagRFP largely redistributed Dvl-GFP to the Gpr124-positive membrane subdomains in both EVL and DEL cells, where the proteins colocalized (Fig. 5D).

We generated a collection of Gpr124 truncation variants with a range of deletions within the 337-amino-acid-long ICD (Fig. 5E) and evaluated their Wnt7/ $\beta$ -catenin functions in vitro (Fig. 5F) and in vivo (Fig. 5G). The Gpr124 and Gpr125 ICDs contain no obvious motifs except for the last four ETTV amino acids that constitute a canonical PDZ-binding motif (PBM), which is also found in a subset of Fz receptors. The ETTV tetrapeptide contributed, but was not essential, for Gpr124 activity. It was similarly dispensable for high-dose Gpr124- or Gpr125-induced planar cell polarity phenotypes (fig. S16) (29). Analysis of increasingly larger C-terminal deletion variants mapped the essential region of the Gpr124 ICD to the interval spanning residues 80 to 213. The activity of the different ICD variants exactly matched their capacity to recruit Dvl-GFP in EVL cells (Fig. 5H) and *Fz<sub>1-10</sub><sup>-/-</sup>* cells (Fig. 5I). This interaction between Gpr124 and Dvl could also be detected with coimmunoprecipitation. In these assays, only Gpr124 variants harboring the 81–213 region interacted with exogenous Dvl-GFP (Fig. 5J) or endogenous DVL2 (fig. S17) in *Fz<sub>1-10</sub><sup>-/-</sup>* cells. This interaction is likely direct because purified recombinant GST-Gpr124-ICD or GST-Gpr124-ICD<sup>ΔETTV</sup> fusion proteins were able to pull down in vitro-translated Dvl-GFP (fig. S18), as previously reported for Gpr125 (29).

To test for the endogenous requirement for Gpr124 ICD interaction with Dvl, we performed gene-disruption experiments in somatic zebrafish embryos by coinjecting *zCas9* mRNA and single-guide RNAs (sgRNAs) targeting the Gpr124 coding sequence immediately upstream, within, or downstream of the Dvl-binding region (ICD residues 80 to 213). Although the injection of four out of five sgRNAs predicted to disrupt Dvl recruitment to Gpr124 generated embryos that lack hindbrain CtAs with a penetrance ranging from 13.6 to 70.8%, none of the seven sgRNAs targeting Gpr124 downstream of the Dvl-binding region generated penetrant brain vascular defects (Fig. 5K).

Taken together, these experiments identify Dvl as a Gpr124 binding partner that could mediate its Wnt7 signaling activities at the plasma membrane. Unlike Fz, Gpr124-mediated recruitment of Dvl at the plasma membrane did not yield a detectable increase in phosphorylated Dvl levels, an early indicator of Wnt signaling activation upstream of  $\beta$ -catenin stabilization (fig. S19) (30, 37). This absence of Gpr124-induced Dvl activation is consistent with the experiments shown in Fig. 4, D to I, that demonstrated Gpr124 to be a conformationally inert Wnt7 signaling mediator.

## Dvl polymers assemble ligand-specific Wnt signalosomes by linking Gpr124 and Fz

Knockdown of *DVL2* by means of small interfering RNA (siRNA) impairs Gpr124/Reck-mediated signaling (fig. S20). Because Dvl is an essential adaptor of Fz, and Gpr124/Reck signaling relies on Fz (fig. S12), such cell-wide loss-of-function approaches are however of limited value to probe Dvl function specifically as a Gpr124 (and not Fz) effector. We therefore used the nanobody strategy described in Fig. 4G to selectively modulate Dvl binding to Gpr124. VhhGFP4-mediated recruitment of Dvl to Gpr124<sup>ΔICD</sup>-GFP, but not to Gpr125<sup>ΔICD</sup>-GFP, was sufficient to partially reverse the *gpr124* mutant vascular phenotype in vivo (Fig. 6A). As additional control, injecting either component alone or substituting VhhGFP4-Dvl with VhhGFP4-RFP did not rescue brain angiogenesis. These experiments reveal that Dvl is sufficient to mediate Gpr124 intracellular functions in Wnt7 signaling.

Gpr124, Reck, and Fz/Lrp5/6 have been reported to form higher-order receptor complexes (24). We reasoned that the Gpr124 ICD might assemble this complex via Dvl. Dvl molecules indeed assemble signalosomes through dynamic polymerization (32–34). Because Dvl physically interacts with both Gpr124 and Fz, Gpr124 and the associated Reck-bound Wnt7 might thus become trapped in dynamic Wnt signalosomes, increasing the local concentration of Wnt7 ligands available for Fz signaling.

Wnt signalosomes are readily detected with light microscopy as large, punctate structures enriched in Dvl that form at or below the plasma membrane (33–35). To determine whether Fz and Gpr124 codistribute in Wnt signalosomes in a Dvl-dependent manner, we first examined the localization of individually expressed Fz-GFP and Gpr124-tagRFP in DEL cells. Fz4 decorated the entire plasma membrane periphery, whereas Gpr124-tagRFP accumulated at cellular contacts (Figs. 5, C and D, and 6B). This differential membrane localization was retained upon Dvl expression (Fig. 6B). Consistent with their Dvl binding capacity, both receptors recruited Dvl-GFP from the cytoplasm (Fig. 5C) to their respective membrane compartments (Fig. 6C). However, when Gpr124-tagRFP and Fz4-GFP were coexpressed, Fz4-GFP quantitatively relocalized, in a Dvl-dependent manner, to the Gpr124-positive intercellular junctions (Fig. 6D). Gpr124-tagRFP and Fz4-GFP colocalized in Wnt signalosome-reminiscent punctate structures that were particularly evident at EVL cell membranes (Fig. 6E and fig. S21).

We used bimolecular fluorescence complementation as an additional assay to test for Dvl-dependent Fz/Gpr124 interaction in DEL cells. Coinjection of Gpr124-VN<sub>155</sub> (I152L) and FzI-VC<sub>155</sub> indeed generated bright junctional signals in a Dvl-dependent manner (Fig. 6F), demonstrating that Fz and Gpr124 indirectly interact via the Dvl scaffold protein. Altogether, these data provide a molecular mechanism for spatial enrichment of Wnt7 within Fz/Lrp5/6 sig-

nalosomes, permitting potentiated and ligand-selective cellular responses (Fig. 6G).

## Discussion

This work provides mechanistic insights onto the Wnt decoding capacities of vertebrate cells. It also demonstrates that the evolutionarily constrained Wnt structure retained enough diversity to allow ligand-specific cellular responses, a property so far thought to require structurally unrelated Frizzled ligands such as Norrin (36).

These structural insights into Wnt evolution and function suggest that additional Wnt decoding modules exist, enabling fine-tuning of cellular behaviors in response to other Wnt or Fz family members. The discrete interaction mode of Fz and Wnt leaves large surfaces made of evolutionary conserved residues available to accommodate additional co-receptors. We therefore propose that Wnt decoding modules might have contributed to shaping the evolution of the Wnt ligand family.

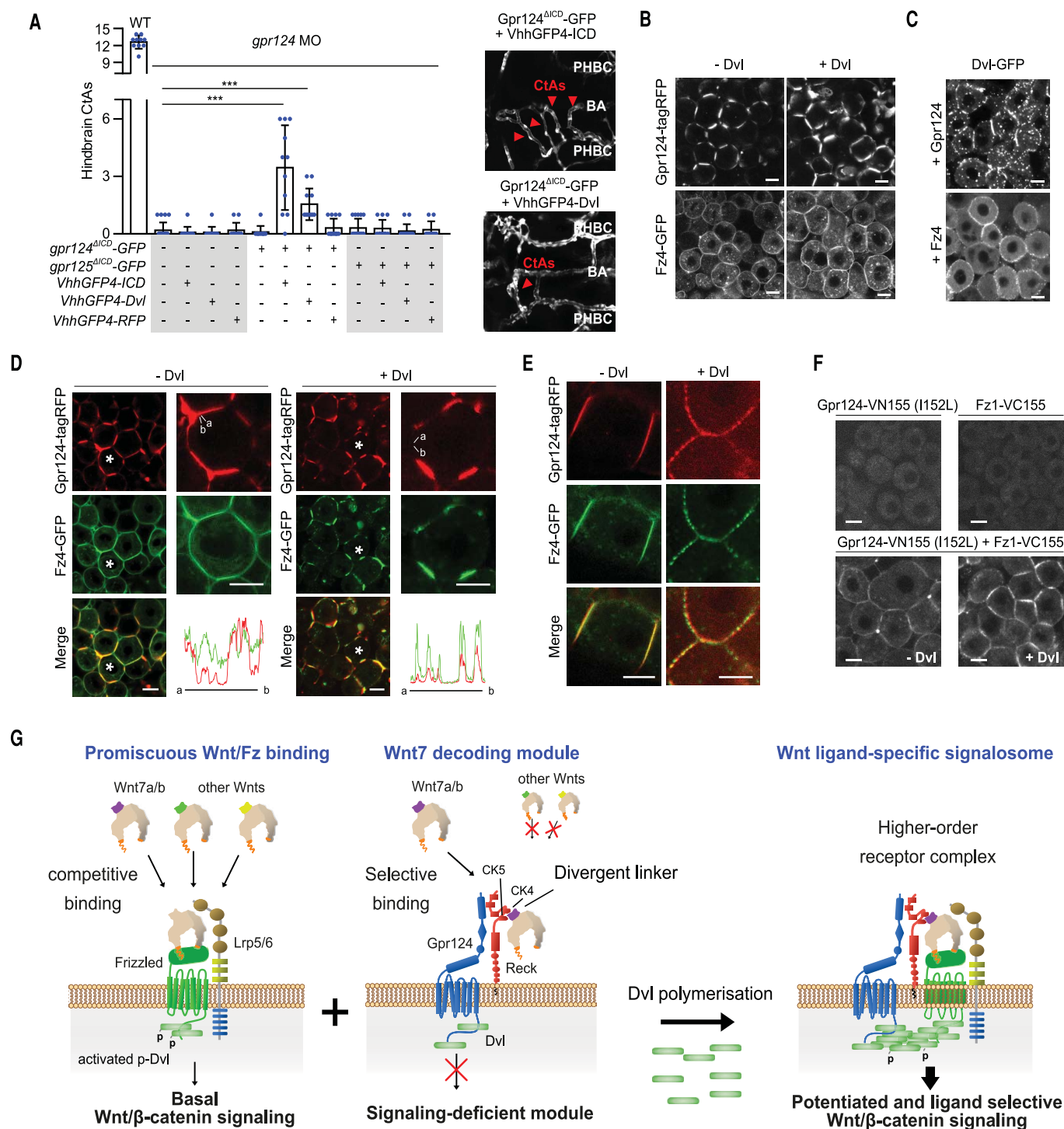
The benefit to promiscuous Wnt/Fz interactions with specificity conferred by accessory proteins rather than monospecific Wnt/Fz interactions might lie in the increased modularity offered by the binary system. A single-component system would be limited to an on-or-off signaling output. The two-component system described here can, context-dependently, achieve cell-autonomous Wnt signaling inhibition or act as a tunable rheostat amplifying the signaling output of specific Wnt ligands.

A salient molecular property of the Wnt7 module is the use of Dvl as a common Gpr124 and Fz adaptor. It is tempting to propose that taking advantage of Fz-associated scaffold proteins such as Dvl could constitute a generic mechanism for Wnt/Fz modifiers. Accordingly, cells recurrently tailor their responses to Wnt by reshaping the molecular composition of the Dvl-associated proteins, including regulatory kinases, E3 ubiquitin ligases, and components of the endocytic machinery (32–34).

Our findings have clinical implications. The pleiotropic functions of Wnt signaling in health and disease make this pathway a conspicuous yet intrinsically challenging therapeutic target. Manipulating the Wnt/ $\beta$ -catenin pathway at the level of its cytosolic or nuclear components harbors high potential for systemic effects with undesirable outcomes across a range of tissues (37). Interventions focused on specific Wnts, Fz receptors, or other signaling components at the cell membrane might in principle be more selective and hence better positioned to lead to clinically viable strategies (4, 38–40). The mechanism uncovered here expands the prospects for specific Wnt-targeted interventions. In particular, by giving molecular insights at the heart of Wnt7-specific signaling, it provides an opportunity for the targeted treatment of human brain disorders with neurovascular involvement, including stroke and brain cancer (41).

## Materials and methods Zebrafish lines

Zebrafish (*Danio rerio*) were maintained at 28°C on a 14 hours light/10 hours dark cycle.



**Fig. 6. Dvl polymers assemble ligand-specific Wnt signalosomes by linking Fz and Gpr124.** (A) Hindbrain CtAs of WT or *gpr124* morphant embryos at 60 hours post fertilization, injected at the one-cell stage with 50 pg of the indicated mRNA. The injected doses of mRNA were reduced compared with Fig. 4H so as to avoid the developmental toxicity of *VhhGFP4-Dvl* mRNA injections. Representative hindbrain vasculature is shown on the right. (B) Intracellular distribution of Fz4-GFP and Gpr124-tagRFP expressed individually or in the presence of Dvl in zebrafish blastula DEL cells. (C) Intracellular distribution of Dvl-GFP coexpressed with Fz4 or Gpr124 in zebrafish blastula DEL cells.

(D) Intracellular distribution of coexpressed Fz4-GFP and Gpr124-tagRFP in the absence or presence of Dvl in zebrafish blastula DEL cells. The cells annotated with an asterisk are magnified at right, and the pixel intensity of the green and red channels along a virtual clockwise path following the cell cortex from a to b is plotted below. (E) Same as (D) in EVL cells. (F) BiFC signals in zebrafish DEL cells expressing Gpr124-VN155 (I152L) and Fz1-VC155 in the presence or absence of Dvl overexpression. (G) Integrated model for Gpr124/Reck-dependent, Wnt7-specific Fz signaling. Scale bars, 10  $\mu$ m. \*\*\* $P < 0.001$ ; data represent mean  $\pm$  SD.



Embryos were obtained and raised under standard conditions in accordance with European and national ethical and animal welfare guidelines (protocol approval number: CEBEA-IBMM-2017-22:65). Staging was performed according to Kimmel *et al.* (42). The transgenic and mutant lines used in this study are *Tg(kdrl:EGFP)<sup>s843</sup>* (43), *Tg(kdrl:HRAS-mCherry)<sup>s896</sup>* (44) and *gpr124<sup>s844</sup>* (22).

### Morpholinos, RNA constructs, and microinjection

Splice-blocking morpholinos targeting *gpr124* (ACTGATATTGATTAACTACCACACA) (22) were purchased from Gene Tools and injected at the one-cell stage at 2 ng. Synthetic mRNAs were transcribed from NotI linearized pCS2 using the mMessage mMachine SP6 Kit (Thermo Fisher Scientific) and injected into one-cell stage zebrafish embryos.

### Somatic gene disruption in zebrafish by CRISPR/Cas9

The *gpr124<sup>ICD</sup>*-targeting sgRNAs constructs (sgRNA 1-12) were obtained by in vitro annealing of the following primers (1-Fo: TAGGTGGCATGCTA-CAAAAGC and 1-Rv: AAACGCTTTTGTAG-CATGCCA; 2-Fo: TAGGTTTGGGTCCGTAG-GTGA and 2-Rv: AAACCTCCACTACGGACC-CAAA; 3-Fo: TAGGGTAATGGTTTGGGTCCGT and 3-Rv: AAACACGGACCAACCATTAC; 4-Fo: TAGGAAGAAGAAGAGTAATGGTT and 4-Rv: AAACAGCACTACTCTTCTTCT; 5-Fo: TAGGAAGCTTGTAGTAGAACCA and 5-Rv: AAAC-TGGTTCTACTACAAGCTT; 6-Fo: TAGGCCA-TTCTGGCCCTCTGAA and 6-Rv: AAACCTCA-GAGGGCCAGAAATGG; 7-Fo: TAGGCCAACG-GGCGGCACAGAG and 7-Rv: AAACCTCTGT-GCCGCCCGTTGG; 8-Fo: TAGGGGCACAGA-GGGGAAGGAC and 8-Rv: AAACGTCTTCC-CCTCTGTGCC; 9-Fo: TAGGTAGTGATGGAG-GTAGCAG and 9-Rv: AAACCTGTACCTC-CATCACTA; 10-Fo: TAGGTGCCTTCGGTAG-CACGTA and 10-Rv: AAACCTACGTGACTACCG-AAGGCA; 11-Fo: TAGGGGTATGAAGACAG-GGTA and 11-Rv: AAACCTACCTGCTCTTCA-TACC; 12-Fo: TAGGTGCGTGGGTATGAAGAGC and 12-Rv: AAACGCTCTTCATACCCACCGA) and cloning into pT7-gRNA (Addgene #46759), as described in (45). sgRNA were transcribed from the BamHI linearized vector pT7-gRNA using the MEGAshortscript T7 kit (Thermo Fisher Scientific). The synthetic Cas9 mRNA was transcribed from the XbaI linearized vector pT3TS-nls-zCas9-nls (Addgene #46757) using the mMessage mMachine T3 Kit (Ambion). sgRNAs 1-12 (50 pg) and nls-zCas9-nls mRNA (150 pg) were injected into one-cell stage zebrafish embryos. The brain vasculature was analyzed and imaged at 48 hpf.

### Whole-mount in situ hybridization (WISH)

Digoxigenin-labeled antisense riboprobes were synthesized with the digoxigenin (DIG) RNA labeling kit (Roche Diagnostics GmbH). Embryos were fixed in 4% paraformaldehyde

overnight at 4°C and whole-mount in situ hybridization was performed using digoxigenin-labeled *myod1* riboprobes as previously described (46, 47).

### Expression plasmid constructs

Wnt, Frizzled, Reck and Gpr124 and their variants used in STF and PLA assays were expressed from the CMV promoter of the pCS2 plasmid except for the following plasmids: pRK5-Lrp5-rhotag, Fz4-GFP (Addgene #42197), Fz1 (Addgene #42253) and Fz5 (Addgene #42267), all kindly provided by J. Nathans. Single-point mutation variants, deletions and chimeras were generated using In-Fusion cloning (ST0345, Takara) and tandem overlapping PCR products. All constructs were confirmed by Sanger sequencing. Throughout this study, we used *Xenopus* Dvl-GFP (Addgene #16788, from the R. Moon laboratory), zebrafish Gpr124 and Gpr125 as well as mouse Reck and untagged Wnt ligands (22). The collection of active human Wnt-V5 ligands was kindly provided by the Xi He laboratory via Addgene (#43807 to #43825) (48). The activity of the V5-tagged ligands was verified by STF assays (fig. S22). Reck deletion variants correspond to the following amino acids of mouse Reck (NP\_057887.2): Reck<sup>ACK1</sup>: 46-93; Reck<sup>ACK2</sup>: 113-150; Reck<sup>ACK3</sup>: 160-206; Reck<sup>ACK4</sup>: 225-272; Reck<sup>ACK5</sup>: 301-347; Reck<sup>ACK1-5</sup>: 46-347; Reck<sup>ACRD</sup>: 352-484 and Reck<sup>AKAZAL</sup>: 636-798. Gpr124 deletion variants correspond to the following amino acids of zebrafish Gpr124 (NP\_001305000.1): Gpr124<sup>AECD</sup>: 68-725 and Gpr124<sup>ATM2-7</sup>: 764-1030. The Gpr124 ICD used in deletion, swapping or fusion constructs corresponds to residues 1058-1367 of the full length protein (28-337 of the ICD), unless otherwise indicated. The ICD of Gpr124 is defined as the segment that follows the last transmembrane segment of the seven pass transmembrane protein as non-ambiguously defined by the InterPro server of EMBL-EBI (based on the Phobius transmembrane topology and signal peptide predictor from the Stockholm Bioinformatics Centre). The ICD of zebrafish Gpr125 (NP\_001289153) and of zebrafish Fz2 (NP\_571215.1) were similarly defined and correspond to the residues 1054-1346 and 526-550, respectively. Gpr124<sup>ECD</sup>-TM<sup>CD27</sup>-GFP was generated by replacing the 7-TM span of Gpr124 (residues 743-1030) by the TM of mouse CD27 (NP\_001028298.1) (IFVTFSSMFLIFVLGAIL). The VhhGFP4-ICD construct was generated by fusing the Gpr124 ICD to the C terminus of VhhGFP4 [kindly provided by Markus Affolter (28)]. Gpr124<sup>ICDFz2ΔKTxxW/ΔETTV</sup> was generated by deleting the KTxxW (KTLHSW) and ETTV motifs within the ICD of zebrafish Fz2. The BiFC constructs were subcloned from pBiFC-VC155 (Addgene #22011) and pBiFC-VN155 (1152L) (Addgene #27097). The HA tag was inserted after residue 22 in murine Reck, residue 47 in zebrafish Gpr124 and residue 22 in murine Fz5. The FLAG tag was inserted after residue 47 in zebrafish Gpr124. We thank the numerous colleagues that facilitated this work by depositing their plasmids at Addgene.

### Cell culture and HEK293(T) mutant cell lines

HEK293T cells were obtained from ATCC (CRL-3216) and the HEK293 STF cell line was kindly provided by J. Nathans. WT and mutant cells were cultured in DMEM/F12 medium (Lonza) supplemented with 10% fetal bovine serum and maintained at 37°C in a humidified incubator equilibrated with 5% CO<sub>2</sub>. *GPR124* and *RECK* were genetically inactivated using CRISPR/Cas9 approaches in HEK293-STF cells and *LRP5*, *LRP6* and *FZ* were inactivated in HEK293T cells. The mutant cell lines were obtained by the iterative CRISPR/Cas9-mediated mutagenesis strategy illustrated in Fig. 1A. CRISPR/Cas9 guide sequences were designed using the <http://crispr.mit.edu> website and were cloned into pSpCas9(BB)-2A-GFP (49). The top 1% of GFP<sup>+</sup> cells was isolated by FACS (AriaIII, BD Biosciences) 48 hours after transfection and distributed in 96-well plates for clonal expansion. In order to facilitate the genetic characterization of the mutant cell lines, clones generating multiple-peak derivative melt curves in high-resolution melt analyses were first counter-selected. For each gene, two independent PCR products flanking the target site (300-1000 bp PCR products centered on the PAM site) were gel purified and characterized by Sanger sequencing on both strands. The absence of WT alleles in the uncloned PCR product was verified and the number of different edited alleles (1, 2 or 3 in hypotriploid HEK293 cells) was deduced from the complexity of the chromatograms (fig. S1). For each gene, the two independent PCR products were then cloned in pCR<sup>TM</sup>-Blunt II-TOPO<sup>®</sup> (Thermo Fisher Scientific) and individual colonies were sequenced until each edited allele was sequenced at least twice from each individual PCR amplicon. The number of analyzed colonies for the *FZ<sub>1-10</sub>*<sup>-/-</sup> cell line is provided in table S1. The sequencing chromatograms form the identified individual clones were then aligned to ensure that the allelic complexity of the uncloned PCR can exhaustively be accounted for by the identified alleles (fig. S1). Finally, to exclude allelic exclusion by PCR amplification bias, the absence of WT alleles in the *FZ<sub>1-10</sub>*<sup>-/-</sup> cell line was verified by whole exome sequencing.

### STF dual luciferase assay

Cells were plated into 96-well plates and transfected after 24 hours in triplicate with Lipofectamine 2000 (Thermo Fisher Scientific). The amount of plasmid DNA transfected per well was optimized for each expression vector as follows: Renilla luciferase (0.5 ng), Wnt ligands (20 ng), Fz receptors (5 ng), Lrp5 (2.5 ng), Gpr124 (10 ng), Reck (5 ng) and Dkk-1 (10 ng) unless otherwise indicated. The total amount of DNA was adjusted to 100 ng per well with the empty pCS2 vector. Dual luciferase assays were performed using the STF cell line or by co-transfecting 20 ng of M50 Super 8x TOPFlash plasmid (Addgene #12456). Cells were harvested in passive lysis buffer (E1980, Promega) and the activities of the Firefly and Renilla luciferases were measured sequentially

using the Dual-Luciferase Reporter Assay system (E1980, Promega) 48 hours post transfection. The competition assays in Fig. 1G and fig. S5 were performed by plating cells as a 1:1:1 mixture at 90% confluency in 96-well plates 24 hours after transfection. Luciferase activity was measured 24 hours after co-culture. For siRNA, cells were plated into 96-well plates and transfected after 24 hours in triplicate using Lipofectamine 2000 (Thermo Fisher Scientific) with 1 pmol (per well) of *Dvl2* or control siRNAs (SASI\_HS01\_00104204 and SIC001, Sigma-Aldrich) together with the indicated plasmids.

### Immunofluorescence and proximity ligation assay

Cells were grown in glass-coated chambers (IBIDI) and transfected after 24 hours with Lipofectamine 2000 (Thermo Fisher Scientific). Cells were fixed with 4% paraformaldehyde for 10 min at room temperature (RT) 48 hours post transfection. For immunofluorescence staining (IF), cells were blocked in 1% BSA-PBS for 30 min before being exposed to primary antibodies for 1 hour at RT. After three PBS washes, cells were incubated with secondary antibodies for 1 hour at RT. For anti-V5 staining of Wnt ligands, cells were additionally washed for 10 min in PBS 0.1% Tween 20 before incubation with the secondary antibody solution. For the proximity ligation assay (PLA) (Sigma-Aldrich), cells were blocked for 30 min at 37°C with the Blocking solution provided by the manufacturer before being incubated with the primary antibodies for 1 hour at RT. Cells were washed three times and incubated with the PLA probes anti-rabbit PLUS and anti-mouse MINUS for 1 hour at 37°C. After two PBS washes, cells were incubated with the Duolink Ligation solution for 30 min at 37°C. After two PBS washes, cells were incubated with the Duolink Amplification solution for 100 min at 37°C. The following antibodies were used: monoclonal mouse anti-V5 (R96025, Thermo Fisher Scientific) at 1:500 for IF and PLA, purified polyclonal rabbit anti-HA (H6908, Sigma-Aldrich) at 1:400 for IF and PLA and anti-mouse Alexa488-conjugated secondary antibody (Thermo Fisher Scientific) at 1:5000. Cells were stained for 2 min with Hoechst diluted to 10  $\mu\text{g ml}^{-1}$  in PBS.

### Western blot, Dot blot, GST pull-down and coimmunoprecipitation

The following antibodies were used with an overnight incubation at 4°C: rabbit anti-DVL2 (1:1000, #3224S, Cell Signaling Technology), rabbit anti-DVL2-phospho T224 (1:1000, ab124941, Abcam), mouse monoclonal anti-V5 (1:1000, R96025, Thermo Fisher Scientific), mouse monoclonal anti- $\beta$ -actin-peroxidase (1:50000, A3854, Sigma-Aldrich), chicken anti-GFP (1:5000, GFP-1010, Aves), and mouse monoclonal anti-FLAG M2 (1:1000, F1804, Sigma-Aldrich).

Dot blot analyses were performed according to manufacturer's protocol with a BioDot SF apparatus (Bio-Rad). Serial dilutions of supernatant were spotted onto a nitrocellulose membrane (GE Healthcare). After drying, the membrane

was incubated with the antibodies as described above. For the coimmunoprecipitations assays, HEK293T were collected 48 hours after transfection from six-well plates and resuspended after two PBS washes in lysis buffer [150 mM NaCl, 25 mM Tris (pH 7.5) and 1% IGEPAL CA-630 (Sigma-Aldrich)] containing EDTA-free protease inhibitor cocktail (Roche) for 30 min at 4°C. After centrifugation at 20000 g for 10 min, the supernatant was incubated with 20  $\mu\text{l}$  of anti-FLAG M2 affinity gel (A2220, Sigma-Aldrich) overnight at 4°C. Beads were washed five times with the lysis buffer and boiled in 2x Laemmli Sample buffer. For GST pull-downs, the coding sequence of Gpr124 ICD and its  $\Delta\text{ETTV}$  variant were fused downstream of GST sequences into pGEX-6PI (GE Healthcare). The GST-fusion proteins were induced in *Escherichia coli* BL21 by exposure to 0.1 mM of IPTG in LB medium at 37°C for 4 hours. Cells were subsequently lysed in a cell disruptor (M110S, Microfluidics) and after centrifugation, the GST-fusion protein from the supernatant were immobilized on Glutathione Sepharose beads (I7-0756-01, GE Healthcare). The Dvl-GFP and GFP constructs were transcribed using the mMESAGE mMACHINE SP6 Transcription Kit (Thermo Fisher Scientific) and in vitro translated using the Rabbit Reticulocyte Lysate System provided by Promega (L4960) in the presence of S<sup>35</sup>-methionine (NEG009T001MC, PerkinElmer). The beads were incubated with the radiolabeled proteins in PBS for 3 hours at 4°C with gentle mixing and then washed five times with cold PBS. The bound complex was resuspended in 2x Laemmli buffer, separated by SDS-page and analyzed by fluorography.

### Microscopy and images processing

Cells and zebrafish embryos were imaged with a LSM710 confocal microscope and images were processed in ImageJ. Images of eye fusion phenotypes were taken on a Leica M165 FC. Brain vasculature renderings were generated using Imaris software (BitPlane).

The following method was used for the quantification of the PLA signal: The in situ PLA signals are seen as bright fluorescent dots of characteristic appearance. The signal lined the cell surface, as expected given the membrane localization of the receptors. For quantification, several images were acquired on LSM710 confocal microscope with a 20x objective. ImageJ was used to generate binary images by global thresholding and to determine the ratio of the total area occupied by the PLA-positive pixels to the DAPI-positive pixels, the latter being indicative of the number of cells in the field of view. The threshold values were manually adjusted to reflect the nuclear localization of DAPI and the focal membrane PLA signals. Identical threshold values were used for all images. On the dot plot, each dot represents the ratio of one image. Each image typically contained 50-100 cells.

### Structural modeling

The structure of *Xenopus* Wnt8a (PDB ID: 4FOA) (13) was used as a starting model for Wnt7a. Missing residues and substitutions were mod-

eled using the program Modeller (50). The modeling strategy included accounting for existing disulphide bridges as observed in XWnt8a. The initial best models were subjected to a conjugate-gradient energy minimization in vacuum with the C $\alpha$  restrained and then freed in a second minimization step. These models were then embedded in a water box and electric neutrality was achieved by adding Na<sup>+</sup> counter ions at 150 mM. The whole system was again energy minimized in 3000 steps. The molecular dynamics simulation was carried out for 0.5 ns with the program NAMD 2.7 at constant temperature (310 K) and constant pressure (1 atm), with periodic boundaries and using CHARMM36 as force field (51). A time step of 2 fs was used to integrate the equations of motion. The short-range interactions were cut at 12 Å and the smooth-particle mesh Ewald method was used to calculate electrostatic interactions. Hydrogen atoms were constrained using the SHAKE algorithm. The resulting model is an average representation of the stable simulation.

### Recombinant Reck-Fc fusions and synthetic peptides

The HA-tagged CK domain of Reck and its individual CK motif deletion variants were fused at the N terminus of the Fc region of human IgG1. The fusion vector was kindly provided by J. Nathans (24). Fc fusion proteins were recovered in serum-free FreeStyle 293 Expression Medium (Thermo Fisher Scientific) from the supernatant of HEK293T cell cultures 72 hours post transfection with Lipofectamine 2000 (Thermo Fisher Scientific). After collection, the supernatants were submitted to Protein G affinity purification (Protein G Sepharose 4 Fast Flow, Sigma-Aldrich). After acidic elution, protein purity was assessed by Coomassie Blue staining (Fig. 3A). Synthetic Wnt linker peptides were obtained from Chinapeptide Co., Ltd at purities of over 90%.

### Isothermal titration calorimetry

ITC titrations were carried out on an Affinity ITC (TA Instruments). Prior to the measurement, Reck-CK-Fc and all Reck-CK-Fc fusions were dialyzed to Tris-NaCl buffer (50 mM Tris pH8, 300 mM NaCl). In each case Wnt-derived peptides were prepared with buffer from the last step of protein dialysis. The samples were filtered and degassed before being examined in the calorimeter and the titrations were performed at 25°C. All the experiments consisted of injection of constant volumes of 2  $\mu\text{L}$  of titrant into the cell (200  $\mu\text{L}$ ) with a stirring rate of 75 rpm. Nominal sample concentrations were between 20  $\mu\text{M}$  and 40  $\mu\text{M}$  in the cell and 400  $\mu\text{M}$  to 1.0 mM in the syringe. Actual sample concentrations were determined after dialysis or buffer exchange by measurement of their absorption at 280 nm or by the BCA method. All data were analyzed using the MicroCal Origin ITC 7.0 and NanoAnalyze software packages.

### Atomic force microscopy

Glass coverslips coated with a thin gold layer were cleaned in an ultraviolet radiation and



ozone (UV-O) cleaner (Jetlight) for 15 min and subsequently immersed overnight in an ethanol solution containing 1 mM 16-mercaptopdodecahexanoic acid and 1-mercapto-1-undecanol at a 1:99 volumetric ratio. Substrates were then rinsed with ethanol, dried with N<sub>2</sub> and added to a solution containing equal volumes of 20 mg ml<sup>-1</sup> N-hydroxysuccinimide (NHS) and 50 mg ml<sup>-1</sup> 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC) for 30 min. The obtained NHS activated surfaces were rinsed with ultrapure water and incubated with 100 µl of a 10 µg ml<sup>-1</sup> Protein G solution for 1 hour at RT. Samples were then washed with washing buffer (3 × 5 min) and further incubated with 100 µl blocking buffer for 1 hour at RT. Finally, 50 µl of a 0.2 µg ml<sup>-1</sup> Reck-CK-Fc solution was added to the substrates for 1 hour, rinsed with washing buffer and subsequently used for AFM experiments. Single-molecule force spectroscopy (SMFS) measurements were performed in PBS buffer at RT using a Nanoscope VIII Multimode AFM (Bruker). Triangular AFM cantilevers (MSCT, Bruker) with silicon nitride tips and a nominal spring constant between 0.01–0.06 N m<sup>-1</sup> were used. Cantilevers were calibrated at the end of each experiment using the thermal noise method (52). Functionalized tips were derivatized using a NHS-PEG<sub>27</sub> Maleimide linker following a protocol described elsewhere (53) to covalently attach the Pep7b linking the cysteine residue added at the C terminus. After functionalization, the cantilevers were washed with PBS (3 × 5 min) and stored in individual wells of a multiwell dish containing 2 ml of PBS per well at 4°C until used in AFM experiments. Force-distance curves were recorded as 32x32 pixels arrays over 1 × 1 µm<sup>2</sup> areas, using an applied force of 250 pN, a contact time of 0.25 s and a constant approach and retraction speed of 1 µm s<sup>-1</sup>. For dynamic force spectroscopy measurements, the retraction speed of the cantilever was varied as follows: 20 nm s<sup>-1</sup>, 100 nm s<sup>-1</sup>, 200 nm s<sup>-1</sup>, 1 µm s<sup>-1</sup>, 2 µm s<sup>-1</sup>, 10 µm s<sup>-1</sup> and 20 µm s<sup>-1</sup>. Typically, at least 2000 force-distance curves were performed for each cantilever at a particular retraction speed. The collected data were analyzed using the Nanoscope Analysis software (Bruker). The retraction segment of each curve was analyzed and unbinding events were considered as specific if they occurred at a distance between 5–50 nm from the contact point. The minimum adhesion force was further used to calculate binding probabilities and build force distribution histograms. To reconstruct the energy landscape of the measured interactions, loading rates were calculated from the force vs time curve, as the slope of the adhesion event before the tip cantilever jumps off to surface. The dependency of the force with the loading rate was then plotted in dynamic force spectroscopy plots.

### Statistical analysis

Statistical analysis was performed using GraphPad software. Data represent mean ± SD. p-values were calculated by the one-way ANOVA (post hoc Dunnett's test) and Student's *t* test for

multiple and single comparisons of normally distributed data (STF and PLA assays) and by the Kruskal–Wallis (post hoc Dunn's test) for multiple comparisons of non-normally distributed data (CtA quantifications); \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001.

### REFERENCES AND NOTES

- C. Y. Logan, R. Nusse, The Wnt signaling pathway in development and disease. *Annu. Rev. Cell Dev. Biol.* **20**, 781–810 (2004). doi: [10.1146/annurev.cellbio.20.010403.113126](#); pmid: [15473860](#)
- R. Nusse, H. E. Varmus, Wnt genes. *Cell* **69**, 1073–1087 (1992). doi: [10.1016/0092-8674\(92\)90630-U](#); pmid: [1617723](#)
- P. Bhanot et al., A new member of the frizzled family from *Drosophila* functions as a Wingless receptor. *Nature* **382**, 225–230 (1996). doi: [10.1038/382225a0](#); pmid: [8717036](#)
- C. Niehrs, The complex world of Wnt receptor signalling. *Nat. Rev. Mol. Cell Biol.* **13**, 767–779 (2012). doi: [10.1038/nrm3470](#); pmid: [23151663](#)
- B. T. MacDonald, X. He, Frizzled and LRP5/6 receptors for Wnt/β-catenin signaling. *Cold Spring Harb. Perspect. Biol.* **4**, a007880 (2012). doi: [10.1101/cshperspect.a007880](#); pmid: [23209147](#)
- J. C. Hsieh, A. Rattner, P. M. Smallwood, J. Nathans, Biochemical characterization of Wnt-frizzled interactions using a soluble, biologically active vertebrate Wnt protein. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3546–3551 (1999). doi: [10.1073/pnas.96.7.3546](#); pmid: [10097073](#)
- A. Kikuchi, H. Yamamoto, S. Kishida, Multiplicity of the interactions of Wnt proteins and their receptors. *Cell. Signal.* **19**, 659–671 (2007). doi: [10.1016/j.cellsig.2006.11.001](#); pmid: [17188462](#)
- C. H. Wu, R. Nusse, Ligand receptor interactions in the Wnt signaling pathway in *Drosophila*. *J. Biol. Chem.* **277**, 41762–41769 (2002). doi: [10.1074/jbc.M207850200](#); pmid: [12205098](#)
- E. J. Rulifson, C. H. Wu, R. Nusse, Pathway specificity by the bifunctional receptor frizzled is determined by affinity for wingless. *Mol. Cell* **6**, 117–126 (2000). doi: [10.1016/S1097-2765\(05\)00018-3](#); pmid: [10949033](#)
- S. L. Holmen, A. Salic, C. R. Zylstra, M. W. Kirschner, B. O. Williams, A novel set of Wnt-Frizzled fusion proteins identifies receptor components that activate β-catenin-dependent signaling. *J. Biol. Chem.* **277**, 34727–34735 (2002). doi: [10.1074/jbc.M204989200](#); pmid: [12121999](#)
- H. Yu, X. Ye, N. Guo, J. Nathans, Frizzled 2 and frizzled 7 function redundantly in convergent extension and closure of the ventricular septum and palate: Evidence for a network of interacting genes. *Development* **139**, 4383–4394 (2012). doi: [10.1242/dev.083352](#); pmid: [23095888](#)
- Y. K. Wang et al., A novel human homologue of the *Drosophila* frizzled wnt receptor gene binds wingless protein and is in the Williams syndrome deletion at 7q11.23. *Hum. Mol. Genet.* **6**, 465–472 (1997). doi: [10.1093/hmg/6.3.465](#); pmid: [9147651](#)
- C. Y. Janda, D. Waghray, A. M. Levin, C. Thomas, K. C. Garcia, Structural basis of Wnt recognition by Frizzled. *Science* **337**, 59–64 (2012). doi: [10.1126/science.1222879](#); pmid: [22653731](#)
- J. M. Stenman et al., Canonical Wnt signaling regulates organ-specific assembly and differentiation of CNS vasculature. *Science* **322**, 1247–1250 (2008). doi: [10.1126/science.1164594](#); pmid: [19023080](#)
- R. Daneman et al., Wnt/β-catenin signaling is required for CNS, but not non-CNS, angiogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 641–646 (2009). doi: [10.1073/pnas.0805165106](#); pmid: [19129494](#)
- S. Lieber et al., Wnt/β-catenin signaling controls development of the blood-brain barrier. *J. Cell Biol.* **183**, 409–417 (2008). doi: [10.1083/jcb.200806024](#); pmid: [18955553](#)
- M. Cullen et al., GPR124, an orphan G protein-coupled receptor, is required for CNS-specific vascularization and establishment of the blood-brain barrier. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 5759–5764 (2011). doi: [10.1073/pnas.1017192108](#); pmid: [21421844](#)
- F. Kuhnert et al., Essential regulation of CNS angiogenesis by the orphan G protein-coupled receptor GPR124. *Science* **330**, 985–989 (2010). doi: [10.1126/science.1196554](#); pmid: [21071672](#)
- K. D. Anderson et al., Angiogenic sprouting into neural tissue requires Gpr124, an orphan G protein-coupled receptor. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 2807–2812 (2011). doi: [10.1073/pnas.1019761108](#); pmid: [21282641](#)
- Y. Zhou, J. Nathans, Gpr124 controls CNS angiogenesis and blood-brain barrier integrity by promoting ligand-specific canonical wnt signaling. *Dev. Cell* **31**, 248–256 (2014). doi: [10.1016/j.devcel.2014.08.018](#); pmid: [25373781](#)
- E. Posokhova et al., GPR124 functions as a WNT7-specific coactivator of canonical β-catenin signaling. *Cell Reports* **10**, 123–130 (2015). doi: [10.1016/j.celrep.2014.12.020](#); pmid: [25558062](#)
- B. Vanhollebeke et al., Tip cell-specific requirement for an atypical Gpr124- and Reck-dependent Wnt/β-catenin pathway during brain angiogenesis. *eLife* **4**, 1–25 (2015). doi: [10.7554/eLife.06489](#); pmid: [26051822](#)
- F. Ulrich et al., Reck enables cerebrovascular development by promoting canonical Wnt signaling. *Development* **143**, 147–159 (2016). doi: [10.1242/dev.123059](#); pmid: [26657775](#)
- C. Cho, P. M. Smallwood, J. Nathans, Reck and Gpr124 Are Essential Receptor Cofactors for Wnt7a/Wnt7b-specific signaling in mammalian CNS angiogenesis and blood-brain barrier regulation. *Neuron* **95**, 1056–1073.e5 (2017). doi: [10.1016/j.neuron.2017.07.031](#); pmid: [28803732](#)
- P. E. Wright, H. J. Dyson, Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**, 18–29 (2015). doi: [10.1038/nrm3920](#); pmid: [25531225](#)
- M. L. H. Chu et al., structural Studies of Wnts and identification of an LRP6 binding site. *Structure* **21**, 1235–1242 (2013). doi: [10.1016/j.str.2013.05.006](#); pmid: [23791946](#)
- N. Bostaille, A. Gauquier, L. Twyffels, B. Vanhollebeke, Molecular insights into Adgr2/Gpr124 and Reck intracellular trafficking. *Biol. Open* **5**, 1874–1881 (2016). doi: [10.1242/bio.021287](#); pmid: [27979830](#)
- S. Harmansa, I. Alborelli, D. Bieli, E. Caussinus, M. Affolter, A nanobody-based toolset to investigate the role of protein localization and dispersal in *Drosophila*. *eLife* **6**, 1–22 (2017). doi: [10.7554/eLife.22549](#); pmid: [28395731](#)
- X. Li et al., Gpr125 modulates Dishevelled distribution and planar cell polarity signaling. *Development* **140**, 3028–3039 (2013). doi: [10.1242/dev.094839](#); pmid: [23821037](#)
- S. Yanagawa, F. van Leeuwen, A. Wodarz, J. Klingensmith, R. Nusse, The Dishevelled protein is modified by wingless signaling in *Drosophila*. *Genes Dev.* **9**, 1087–1097 (1995). doi: [10.1101/gad.9.9.1087](#); pmid: [7744250](#)
- X. Huang et al., Phosphorylation of Dishevelled by protein kinase RIKP4 regulates Wnt signaling. *Science* **339**, 1441–1445 (2013). doi: [10.1126/science.1232253](#); pmid: [23371553](#)
- M. Gammons, M. Bienz, Multiprotein complexes governing Wnt signal transduction. *Curr. Opin. Cell Biol.* **51**, 42–49 (2018). doi: [10.1016/j.cob.2017.10.008](#); pmid: [29153704](#)
- M. Bienz, Signalosome assembly by domains undergoing dynamic head-to-tail polymerization. *Trends Biochem. Sci.* **39**, 487–495 (2014). doi: [10.1016/j.tibs.2014.08.006](#); pmid: [25239056](#)
- J. Bilic et al., Wnt induces LRP6 signalosomes and promotes dishevelled-dependent LRP6 phosphorylation. *Science* **316**, 1619–1622 (2007). doi: [10.1126/science.1137065](#); pmid: [17569865](#)
- M. V. Gammons, M. Renko, C. M. Johnson, T. J. Rutherford, M. Bienz, Wnt signalosome assembly by DEP domain swapping of Dishevelled. *Mol. Cell* **64**, 92–104 (2016). doi: [10.1016/j.molcel.2016.08.026](#); pmid: [27692984](#)
- M. B. Lai et al., TSPAN12 is a Norrin Co-receptor that amplifies Frizzled4 ligand selectivity and signaling. *Cell Reports* **19**, 2809–2822 (2017). doi: [10.1016/j.celrep.2017.06.004](#); pmid: [28658627](#)
- Z. F. Zimmerman, R. T. Moon, A. J. Chien, Targeting Wnt pathways in disease. *Cold Spring Harb. Perspect. Biol.* **4**, a008086 (2012). doi: [10.1101/cshperspect.a008086](#); pmid: [23001988](#)
- A. Koval, V. L. Katanaev, Platforms for high-throughput screening of Wnt/Frizzled antagonists. *Drug Discov. Today* **17**, 1316–1322 (2012). doi: [10.1016/j.drudis.2012.07.007](#); pmid: [22819927](#)
- H. Clevers, R. Nusse, Wnt/β-catenin signaling and disease. *Cell* **149**, 1192–1205 (2012). doi: [10.1016/j.cell.2012.05.012](#); pmid: [22682243](#)
- E. Driehuis, H. Clevers, WNT signalling events near the cell membrane and their pharmacological targeting for the treatment of cancer. *Br. J. Pharmacol.* **174**, 4547–4563 (2017). doi: [10.1111/bph.13758](#); pmid: [28244067](#)
- J. Chang et al., Gpr124 is essential for blood-brain barrier integrity in central nervous system disease. *Nat. Med.* **23**, 450–460 (2017). doi: [10.1038/nm.4309](#); pmid: [28288111](#)
- C. B. Kimmel, W. W. Ballard, S. R. Kimmel, B. Ullmann, T. F. Schilling, Stages of embryonic development of the

- zebrafish. *Dev. Dyn.* **203**, 253–310 (1995). doi: [10.1002/aja.1002030302](https://doi.org/10.1002/aja.1002030302); pmid: [8589427](https://pubmed.ncbi.nlm.nih.gov/8589427/)
43. S. W. Jin, D. Beis, T. Mitchell, J. N. Chen, D. Y. R. Stainier, Cellular and molecular analyses of vascular tube and lumen formation in zebrafish. *Development* **132**, 5199–5209 (2005). doi: [10.1242/dev.02087](https://doi.org/10.1242/dev.02087); pmid: [16251212](https://pubmed.ncbi.nlm.nih.gov/16251212/)
  44. N. C. Chi *et al.*, Foxn4 directly regulates tbx2b expression and atrioventricular canal formation. *Genes Dev.* **22**, 734–739 (2008). doi: [10.1101/gad.1629408](https://doi.org/10.1101/gad.1629408); pmid: [18347092](https://pubmed.ncbi.nlm.nih.gov/18347092/)
  45. L. E. Jao, S. R. Wente, W. Chen, Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13904–13909 (2013). doi: [10.1073/pnas.1308335110](https://doi.org/10.1073/pnas.1308335110); pmid: [23918387](https://pubmed.ncbi.nlm.nih.gov/23918387/)
  46. E. S. Weinberg *et al.*, Developmental regulation of zebrafish MyoD in wild-type, no tail and spadetail embryos. *Development* **122**, 271–280 (1996). pmid: [8565839](https://pubmed.ncbi.nlm.nih.gov/8565839/)
  47. C. Thisse, B. Thisse, High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat. Protoc.* **3**, 59–69 (2008). doi: [10.1038/nprot.2007.514](https://doi.org/10.1038/nprot.2007.514); pmid: [18193022](https://pubmed.ncbi.nlm.nih.gov/18193022/)
  48. B. T. MacDonald *et al.*, Disulfide bond requirements for active Wnt ligands. *J. Biol. Chem.* **289**, 18122–18136 (2014). doi: [10.1074/jbc.M114.575027](https://doi.org/10.1074/jbc.M114.575027); pmid: [24841207](https://pubmed.ncbi.nlm.nih.gov/24841207/)
  49. F. A. Ran *et al.*, Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013). doi: [10.1038/nprot.2013.143](https://doi.org/10.1038/nprot.2013.143); pmid: [24157548](https://pubmed.ncbi.nlm.nih.gov/24157548/)
  50. B. Webb, A. Sali, *Current Protocols in Bioinformatics* (John Wiley & Sons, 2016), vol. 54, p. 5.6.1-5.6.37.
  51. J. C. Phillips *et al.*, Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005). doi: [10.1002/jcc.20289](https://doi.org/10.1002/jcc.20289); pmid: [16222654](https://pubmed.ncbi.nlm.nih.gov/16222654/)
  52. H. J. Butt, M. Jaschke, Calculation of thermal noise in atomic force microscopy. *Nanotechnology* **6**, 1–7 (1995). doi: [10.1088/0957-4484/6/1/001](https://doi.org/10.1088/0957-4484/6/1/001)
  53. L. Wildling *et al.*, Probing binding pocket of serotonin transporter by single molecular force spectroscopy on living cells. *J. Biol. Chem.* **287**, 105–113 (2012). doi: [10.1074/jbc.M111.304873](https://doi.org/10.1074/jbc.M111.304873); pmid: [22033932](https://pubmed.ncbi.nlm.nih.gov/22033932/)
  54. R. W. Friddle, A. Noy, J. J. De Yoreo, Interpreting the widespread nonlinear force spectra of intermolecular bonds. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 13573–13578 (2012). doi: [10.1073/pnas.1202946109](https://doi.org/10.1073/pnas.1202946109); pmid: [22869712](https://pubmed.ncbi.nlm.nih.gov/22869712/)

#### ACKNOWLEDGMENTS

We thank J. Nathans (Johns Hopkins), R. Nusse (Stanford), and V. L. Katanaev (UNIL) for advice and fruitful discussions. We thank E. Dupont, S. Rosar, J. Marchal, V. Micha, S. Denanglaire, J. Declercq, L. Sanderson, M. Boeckstaens, N. De Henau, and L. Twyffels

(ULB) for their help. **Funding:** M.E. and N.B. are FRIA fellows from Fonds de la Recherche Scientifique–FNRS. P.C. and D.A. are a postdoctoral researcher and research associate of the FRS–FNRS, respectively. Work in the B.V. laboratory is supported by FNRS (MIS F.4543.15), the Concerted Research Action (ARC), the Fondation ULB, the Queen Elisabeth Medical Foundation (Q.E.M.F.), and FRFS-WELBIO (CR-2017S-05). CMMI is supported by the European Regional Development Fund and the Walloon Region. **Author contributions:** All authors performed research and/or analyzed data. All authors discussed results and edited the manuscript. B.V. designed the research and wrote the manuscript. **Competing interests:** The ULB (B.V.) has filed a patent on the neurological applications of the Gpr124/Reck-mediated signaling mechanism. **Data and materials availability:** All data are available in the main text or the supplementary materials. Correspondence and requests for materials should be addressed to B.V.

#### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/361/6403/eaat1178/suppl/DC1](http://www.sciencemag.org/content/361/6403/eaat1178/suppl/DC1)  
Figs. S1 to S22  
Table S1

30 January 2018; accepted 26 June 2018  
Published online 19 July 2018  
[10.1126/science.aat1178](https://doi.org/10.1126/science.aat1178)



## REPORT

## ORGANIC CHEMISTRY

# Stereodivergent synthesis of 1,4-dicarbonyls by traceless charge-accelerated sulfonium rearrangement

Dainis Kaldre\*, Immo Klose\*, Nuno Maulide†

The chemistry of the carbonyl group is essential to modern organic synthesis. The preparation of substituted, enantioenriched 1,3- or 1,5-dicarbonyls is well developed, as their disconnection naturally follows from the intrinsic polarity of the carbonyl group. By contrast, a general enantioselective access to quaternary stereocenters in acyclic 1,4-dicarbonyl systems remains an unresolved problem, despite the tremendous importance of 2,3-substituted 1,4-dicarbonyl motifs in natural products and drug scaffolds. Here we present a broad enantioselective and stereodivergent strategy to access acyclic, polysubstituted 1,4-dicarbonyls via acid-catalyzed [3,3]-sulfonium rearrangement starting from vinyl sulfoxides and ynamides. The stereochemistry at sulfur governs the absolute sense of chiral induction, whereas the double bond geometry dictates the relative configuration of the final products.

Much of organic synthesis revolves around the chemistry of the carbonyl group. Conjugations of more than one carbonyl function are given special attention and referred to by the relative disposition of the two moieties (e.g., 1,3; 1,4; or 1,5), along with strategies for their direct preparation. Highly efficient approaches exist for the preparation of substituted, enantioenriched 1,3- or 1,5-dicarbonyls, as their disconnection naturally follows from the intrinsic polarity of the carbonyl group. By contrast, the 1,4-dicarbonyl pattern remains challenging to access, even though 2,3-substituted 1,4-dicarbonyl motifs are commonly found in numerous natural products (1) and drug scaffolds (2, 3) and are key synthons for many named reactions in organic chemistry (4, 5) (Fig. 1A). Most of the current methods for the synthesis of these motifs involve direct formation of the central, C2–C3 bond via oxidative coupling (1, 6–8) or Umpolung strategies (9, 10) involving enolate alkylation (11) (Fig. 1B). However, whereas the first family of methods fails to properly address stereoselectivity, the second one is limited in structural flexibility or requires multistep preparation of two chiral starting materials while being only moderately diastereoselective. Rearrangements have been used to access 1,4-dicarbonyls (12, 13), but these methods are strongly limited to certain molecular scaffolds. A general enantioselective route to quaternary stereocenters in acyclic

1,4-dicarbonyl systems remains an open problem (1, 14–16).

Here we report a broad strategy to access acyclic, polysubstituted 1,4-dicarbonyls via charge-accelerated sulfonium rearrangement (17–19). Our approach

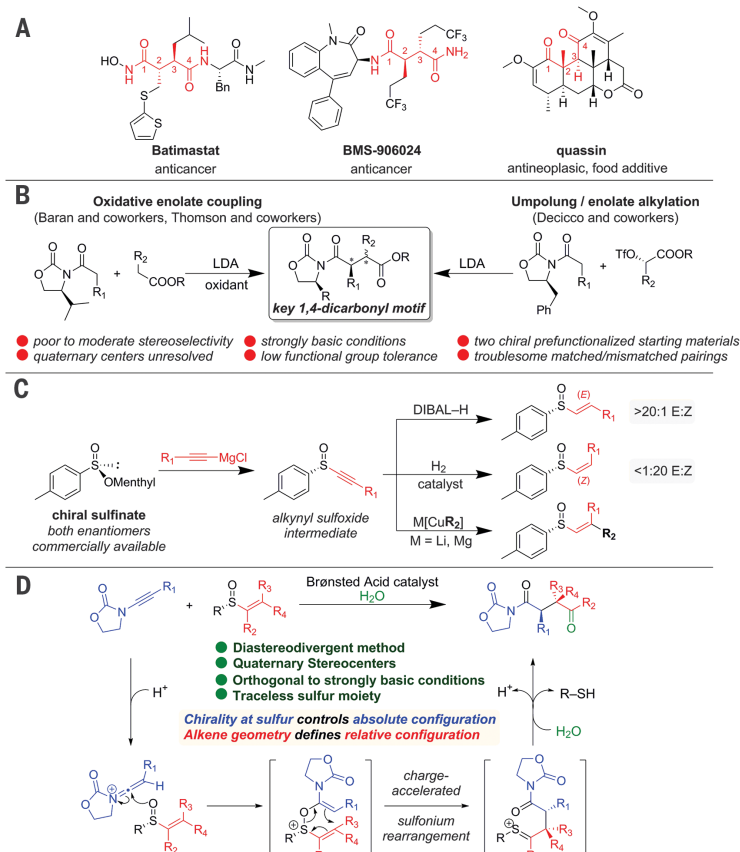
uses highly enantioenriched alkenylsulfoxides—readily available substrates in two steps from commercially available menthyl sulfinates (Fig. 1C) (20–23). In combination with ynamides (24) and a Brønsted acid catalyst, these undergo a [3,3]-sulfonium rearrangement to form a thionium intermediate that is hydrolyzed in situ to give the respective aldehydes or ketones in a traceless manner (Fig. 1D). This catalytic approach allows the preparation of tertiary and quaternary centers to access each and every diastereomer and enantiomer of the 1,4-dicarbonyl products at will and with high stereopurity.

Our investigation started with the use of (*E*)-vinyl sulfoxide **1a** (22, 23), which enabled the selective formation of *syn*-2,3-disubstituted 1,4-dicarbonyls (Fig. 2A). During optimization studies, both the addition of water to the reaction mixture and the use of oxazolidinone-derived ynamides afforded superior results (see supplementary materials for more details). Several aliphatic ynamides afforded the desired aldehydes in good yields and high diastereomeric ratio (d.r.), and the stereoselectivity was further improved when aromatic ynamides were used (**2d**, **2e**). Numerous base-sensitive functional groups such as esters (**2f**), nitriles (**2g**), imides (**2h**), and ketones (**2i**), as well as primary chlorides, aldehydes, and

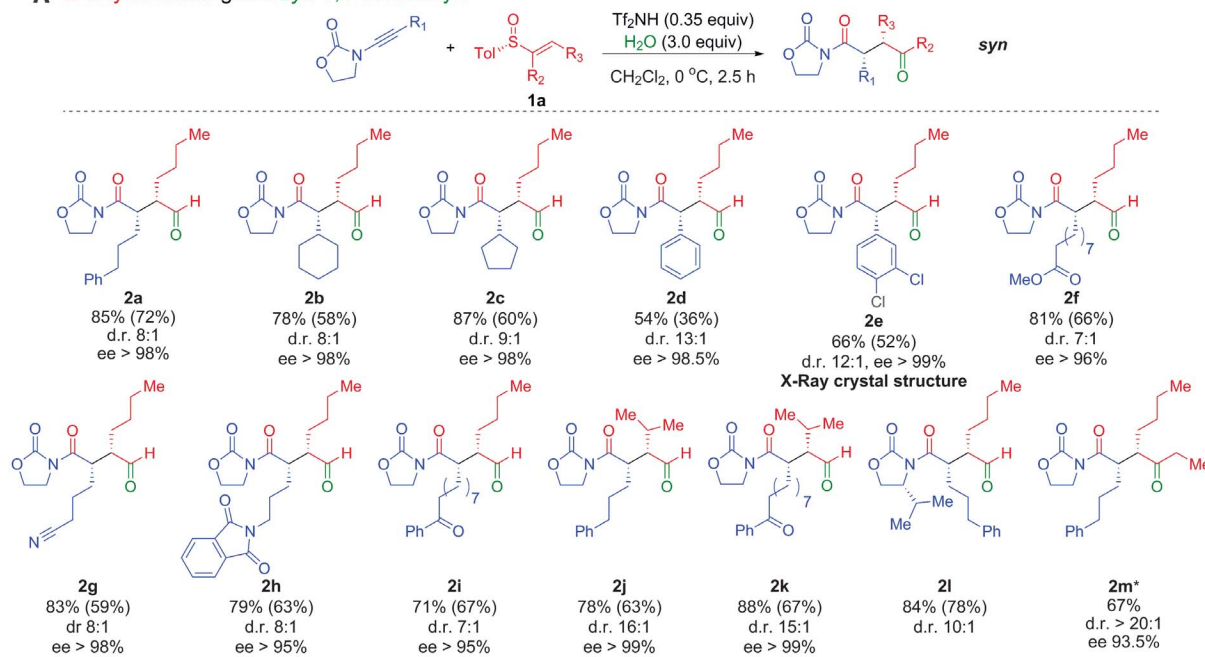
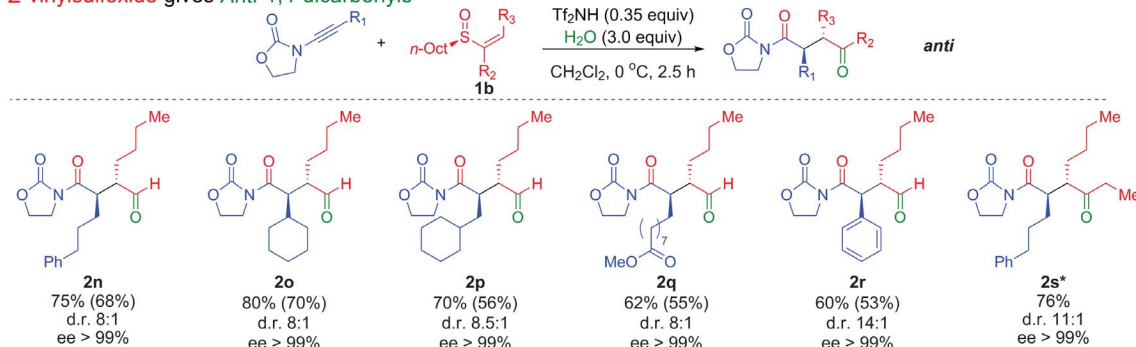
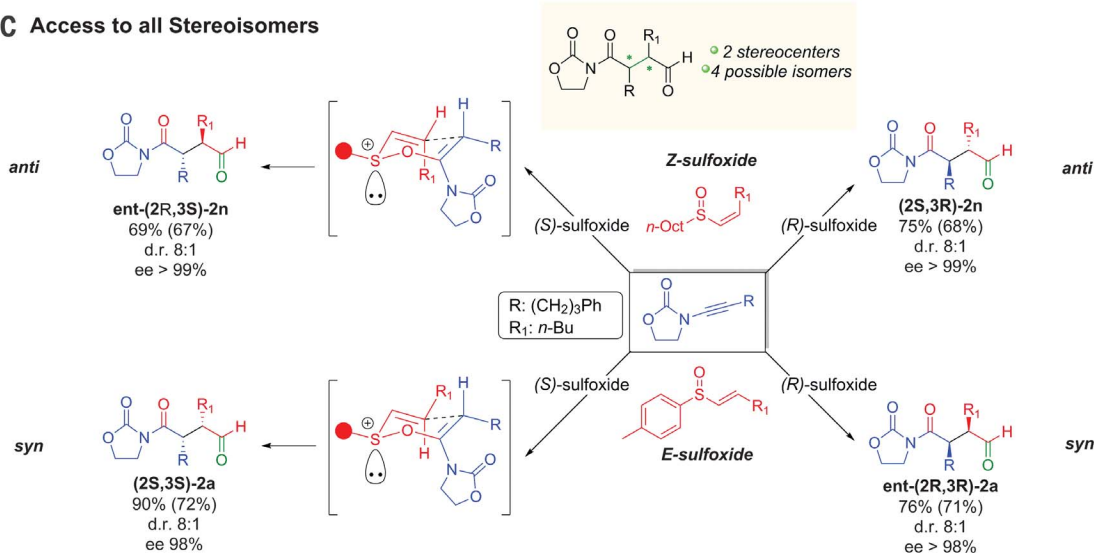
Institute of Organic Chemistry, University of Vienna, Währinger Strasse 38, 1090 Vienna, Austria

\*These authors contributed equally to this work.

†Corresponding author. Email: nuno.maulide@univie.ac.at



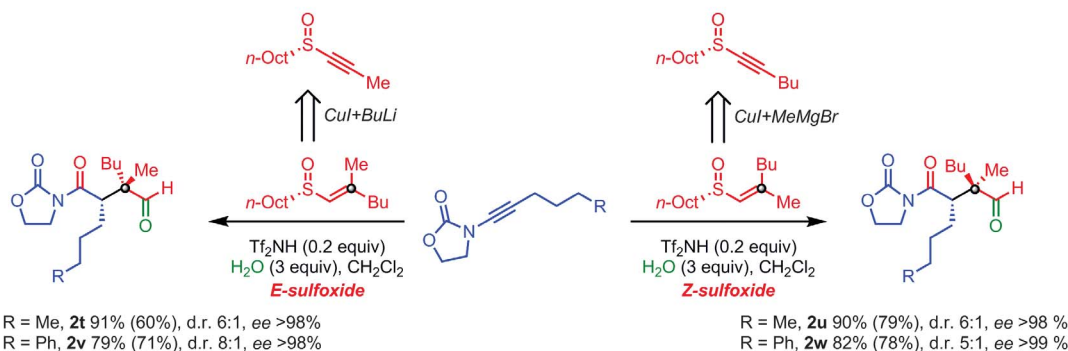
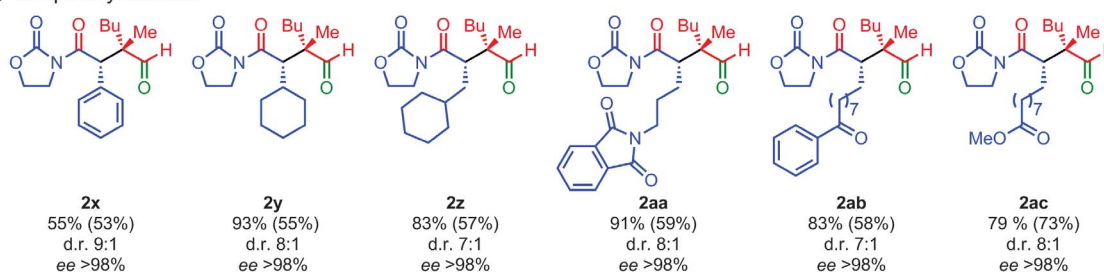
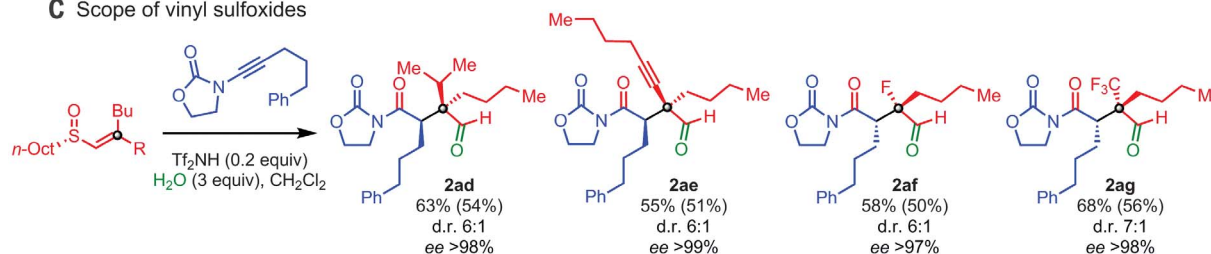
**Fig. 1. Relevance and synthesis of 1,4-dicarbonyls.** (A) 1,4-dicarbonyl motifs in bioactive substances. (B) Current strategies and common limitations. (C) Highly modular and stereoselective synthesis of vinyl sulfoxides. (D) Stereodivergent approach for the enantio- and diastereoselective synthesis of 1,4-dicarbonyls.

**A** *E*-vinylsulfoxide gives *Syn*-1,4-dicarbonyls**B** *Z*-vinylsulfoxide gives *Anti*-1,4-dicarbonyls**C** Access to all Stereoisomers

**Fig. 2. Substrate scope.** (A) *syn*-1,4-dicarbonyls. (B) *anti*-1,4-dicarbonyls. (C) Access to all possible stereoisomers of a 1,4-dicarbonyl product. Unless otherwise indicated, reactions were run on 0.1- to 0.2-mmol scale. Yields were determined by  $^1\text{H}$ -NMR (nuclear magnetic resonance) using an internal

standard (isolated yields shown in parentheses). Diastereomeric ratios were determined by  $^1\text{H}$ -NMR analysis of the crude product. Enantiomeric excess (ee) determined via high-performance liquid chromatography (HPLC). \* Isolated yield. Modified conditions used, see supplementary materials for details.



**A** Vinyl sulfoxide geometry dictates *relative* and *absolute* configuration for *all-carbon quaternary stereocenters***B** Scope of ynamides**C** Scope of vinyl sulfoxides

**Fig. 3. Synthesis of all-carbon quaternary products.** (A) Preparation of quaternary centers. (B) Scope of ynamides. (C) Scope of vinyl sulfoxides, including fluorine and trifluoromethyl derivatives. Unless otherwise indicated,

reactions were run on 0.1- to 0.2-mmol scale. Yields were determined by  $^1\text{H-NMR}$  using an internal standard. Isolated yields in parentheses. d.r. ratios were determined by  $^1\text{H-NMR}$  analysis of the crude product.

alkynes (see supplementary materials), were all well-tolerated. Sterically more demanding substituents on the sulfoxide greatly enhanced the stereoselectivity, delivering the desired products with high diastereoselectivity (**2j** and **2k**). When using ynamides with a chiral auxiliary, we observed matched-mismatched pairings leading to diastereomeric ratios of 10:1 (matched, compare **2l**) and 3:1 (mismatched, compare supplementary materials) respectively. The products were isolated as free aldehydes ready for further functionalization (see below), and ketones (**2m**) could also be accessed using an  $\alpha$ -substituted sulfoxide ( $R_2 \neq \text{H}$ ).

Conversely, the use of (*Z*)-sulfoxides resulted in the formation of *anti*-1,4-dicarbonyl products (Fig. 2B). In this case, the use of alkyl vinyl sulfoxides led to increased yields and stereoselectivities (see supplementary materials for details). Consistently high diastereo- and enantioselectivities were observed. Thus, all four possible isomers (**2a**, **2n**, and their respective enantiomers) of a 1,4-dicarbonyl product are accessible by this stereodivergent approach, in

high yield and stereoselectivity, by simply switching between double bond geometry and sulfoxide stereoisomers, as demonstrated in Fig. 2C (25–29).

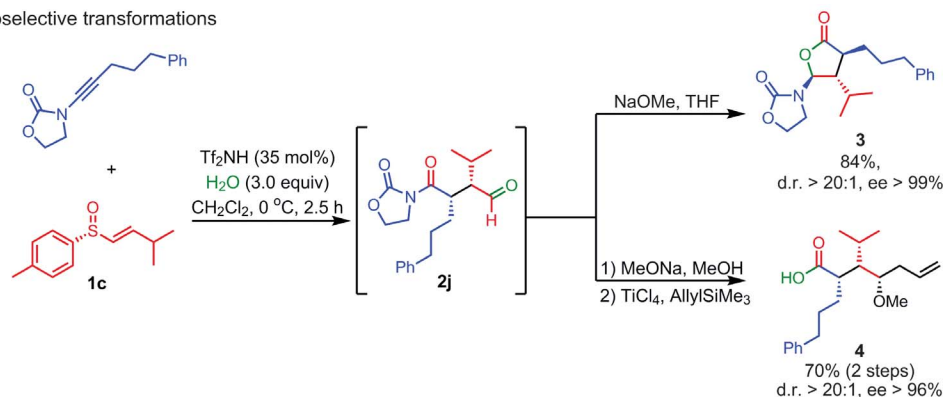
In our stereochemical model, the charge-accelerated sulfonium rearrangement is consistent with a chair-like transition state. The (*E*)-vinyl sulfoxides bias all substituents into a pseudoequatorial orientation, whereas their (*Z*)-counterparts mandate that the  $R_1$  substituent (Fig. 2C) occupies a pseudoaxial orientation directing the substituents to an *anti*-relationship in a final 1,4-dicarbonyl structure. The stereochemistry at sulfur governs the absolute sense of chiral induction, whereas the double bond geometry in turn dictates the relative stereochemistry of the final products.

Having successfully demonstrated diastereo- and enantiodivergence, we turned our attention to  $\beta,\beta$ -disubstituted alkenylsulfoxides to access all-carbon quaternary stereocenters in a stereoselective fashion. As shown in Fig. 3, A and B, the easily controlled sulfoxide double bond geometry correspondingly induces formation of the quaternary carbon stereocenter with high diastereoselectivity and perfect enantiocontrol.

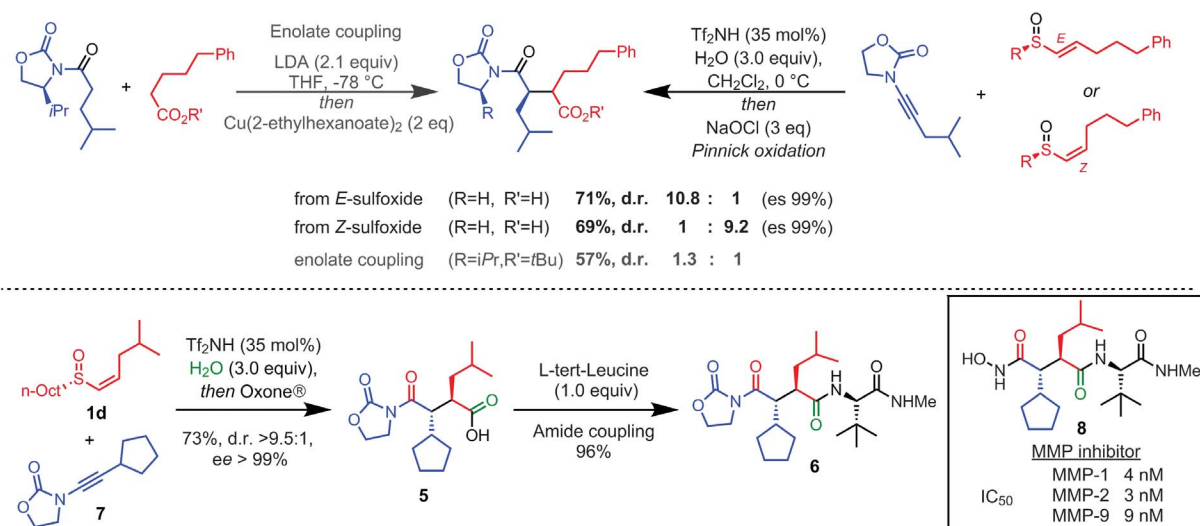
Exploring this approach further, we were able to install isopropyl (**2ad**), alkynyl (**2ae**), fluoro (**2af**), and trifluoromethyl substituents (**2ag**) on the newly formed quaternary center, all of which usually would require their own synthetic strategy (Fig. 3C) (30–33).

Enantioenriched polysubstituted 1,4-dicarbonyls are important intermediates in synthesis. For example, diastereoselective nucleophilic additions to the aldehyde moiety easily afforded trisubstituted lactone **3** (Fig. 4A) (34) or the two-step protocol afforded  $\gamma$ -oxobutyric acid building block **4** (35). A direct comparison of our method to state-of-the-art enolate coupling highlights its advantages, in that both stereoisomers become available at will and in high purity under comparably mild, catalytic conditions. Several highly potent MMP inhibitors bear a 2,3-disubstituted 1,4-dicarbonyl backbone, and the simple access to fully functionalized succinate **6** with excellent stereocontrol (Fig. 4B) is representative of the synthetic value of this method (2, 36). That the fully annotated precursor **5** is elaborated in a single-step from the unconventional building blocks **1d** and **7** is a hallmark of this strategy.

## A Diastereoselective transformations



## B Direct stereoselective access to important succinate building blocks



**Fig. 4. Applications.** (A) Diastereoselective transformations. (B) Direct stereoselective access to succinate building blocks. Yields refer to isolated material. d.r. ratios were determined by  $^1\text{H}$ -NMR analysis of the crude product.

## REFERENCES AND NOTES

- M. P. DeMartino, K. Chen, P. S. Baran, *J. Am. Chem. Soc.* **130**, 11546–11560 (2008).
- M. Whittaker, C. D. Floyd, P. Brown, A. J. H. Gearing, *Chem. Rev.* **99**, 2735–2776 (1999).
- T. Fujisawa et al., *Bioorg. Med. Chem.* **10**, 2569–2581 (2002).
- C. Paal, *Ber. Dtsch. Chem. Ges.* **17**, 2756–2767 (1884).
- L. Knorr, *Ber. Dtsch. Chem. Ges.* **17**, 2863–2870 (1884).
- P. S. Baran, M. P. DeMartino, *Angew. Chem. Int. Ed.* **45**, 7083–7086 (2006).
- E. E. Robinson, R. J. Thomson, *J. Am. Chem. Soc.* **140**, 1956–1965 (2018).
- N. Kise, K. Tokioka, Y. Aoyama, Y. Matsumura, *J. Org. Chem.* **60**, 1100–1101 (1996).
- S. R. Yetra, A. Patra, A. T. Biju, *Synthesis* **47**, 1357–1378 (2015).
- D. Kaiser, C. J. Teskey, P. Adler, N. Maulide, *J. Am. Chem. Soc.* **139**, 16040–16043 (2017).
- C. P. Decicco, D. J. Nelson, R. L. Corbett, J. C. Dreabitt, *J. Org. Chem.* **60**, 4782–4785 (1996).
- L. M. Pratt et al., *Synlett* **1998**, 531–533 (1998).
- S. Huang, L. Kötzer, C. K. De, B. List, *J. Am. Chem. Soc.* **137**, 3446–3449 (2015).
- Y. Liu, S.-J. Han, W.-B. Liu, B. M. Stoltz, *Acc. Chem. Res.* **48**, 740–751 (2015).
- E. A. Ildi, C. E. Stivala, A. Zakarian, *Chem. Soc. Rev.* **38**, 3133–3148 (2009).
- Y. Minko, M. Pasco, L. Lercher, M. Botoshansky, I. Marek, *Nature* **490**, 522–526 (2012).
- X. Huang, S. Klimczyk, N. Maulide, *Synthesis* **2**, 175–183 (2012).
- T. Yanagi et al., *J. Am. Chem. Soc.* **138**, 14582–14585 (2016).
- Z. He et al., *Angew. Chem. Int. Ed.* **57**, 5759–5764 (2018).
- D. Kaldre et al., *Angew. Chem. Int. Ed.* **56**, 2212–2215 (2017).
- J. P. Marino, M. Neisser, *J. Am. Chem. Soc.* **103**, 7687–7689 (1981).
- H. Kosugi, M. Kitaoka, K. Tagami, A. Takahashi, H. Uda, *J. Org. Chem.* **52**, 1078–1082 (1987).
- H. B. Kagan, *Asymmetric synthesis of chiral sulfoxides*, in *Organosulfur Chemistry in Asymmetric Synthesis*, T. Toru, C. Bolm, Eds. (Wiley-VCH, Weinheim, Germany, 2008), pp. 1–29.
- K. A. DeKorver et al., *Chem. Rev.* **110**, 5064–5106 (2010).
- M. T. Oliveira, M. Luparia, D. Audisio, N. Maulide, *Angew. Chem. Int. Ed.* **52**, 13149–13152 (2013).
- M. Luparia et al., *Angew. Chem. Int. Ed.* **50**, 12631–12635 (2011).
- S. Krautwald, D. Sarlah, M. A. Schafröth, E. M. Carreira, *Science* **340**, 1065–1068 (2013).
- X. Huo, R. He, X. Zhang, W. Zhang, *J. Am. Chem. Soc.* **138**, 11093–11096 (2016).
- X. Jiang, J. J. Beiger, J. F. Hartwig, *J. Am. Chem. Soc.* **139**, 87–90 (2017).
- Y. Zhu et al., *Chem. Rev.* **118**, 3887–3964 (2018).
- J. Saadi, H. Wennemers, *Nat. Chem.* **8**, 276–280 (2016).
- D. A. Nagib, M. E. Scott, D. W. C. MacMillan, *J. Am. Chem. Soc.* **131**, 10875–10877 (2009).
- N. Früh, A. Togni, *Angew. Chem. Int. Ed.* **53**, 10813–10816 (2014).
- B. Mao, M. Fañanás-Mastral, B. L. Feringa, *Chem. Rev.* **117**, 10502–10566 (2017).
- A. van Oeveren, B. L. Feringa, *Tetrahedron Lett.* **35**, 8437–8440 (1994).
- A. V. Gavai et al., *ACS Med. Chem. Lett.* **6**, 523–527 (2015).

## ACKNOWLEDGMENTS

I. A. Roller (University of Vienna) is acknowledged for assistance with crystallographic structure determination and E. Macoratti (University of Vienna) for HPLC analysis. **Funding:** We are grateful to the ERC (CoG VINCAT), the FWF (P30226), and the DFG (grant MA 4861/4-2) for financial support of this research. Generous continued support by the University of Vienna is acknowledged. **Author contributions:** N.M. conceived the project; D.K. and I.K. carried out the experiments; N.M., D.K., I.K. wrote the manuscript. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** X-ray structural data are available free of charge from the Cambridge Crystallographic Data Centre under CCDC 1828062. Other characterization data, optimization tables, and additional substrates are in the supplementary materials. Requests for materials should be addressed to N.M.

## SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6403/664/suppl/DC1  
 Materials and Methods  
 Table S1 to S6  
 Figs. S1 to S8  
 NMR Spectra  
 HPLC Spectra  
 References (37–52)

15 March 2018; accepted 12 June 2018  
 10.1126/science.aat5883



## ORGANIC CHEMISTRY

# Selective functionalization of methane, ethane, and higher alkanes by cerium photocatalysis

Anhua Hu\*, Jing-Jing Guo\*, Hui Pan, Zhiwei Zuo†

With the recent soaring production of natural gas, the use of methane and other light hydrocarbon feedstocks as starting materials in synthetic transformations is becoming increasingly economically attractive, although it remains chemically challenging. We report the development of photocatalytic C–H amination, alkylation, and arylation of methane, ethane, and higher alkanes under visible light irradiation at ambient temperature. High catalytic efficiency (turnover numbers up to 2900 for methane and 9700 for ethane) and selectivity were achieved using abundant, inexpensive cerium salts as photocatalysts. Ligand-to-metal charge transfer excitation generated alkoxy radicals from simple alcohols that in turn acted as hydrogen atom transfer catalysts. The mixed-phase gas/liquid reaction was adapted to continuous flow, enabling the efficient use of gaseous feedstocks in scalable photocatalytic transformations.

Methane and other gaseous alkanes ( $C_2$  to  $C_4$ ) have traditionally been viewed more as fuels than as economical chemical feedstocks. However, the recent discovery of huge volumes of unconventional reservoirs and the ensuing soaring production of natural gas have made these gaseous hydrocarbons economically attractive and strategically important basic raw materials (1, 2). With the added economic and ecological benefits of reducing transportation costs and emissions, the direct transformation of gaseous hydrocarbons into value-added liquid commodity chemicals via the use of innovative homogeneous catalysis processes has received considerable attention in recent years (1, 3–8).

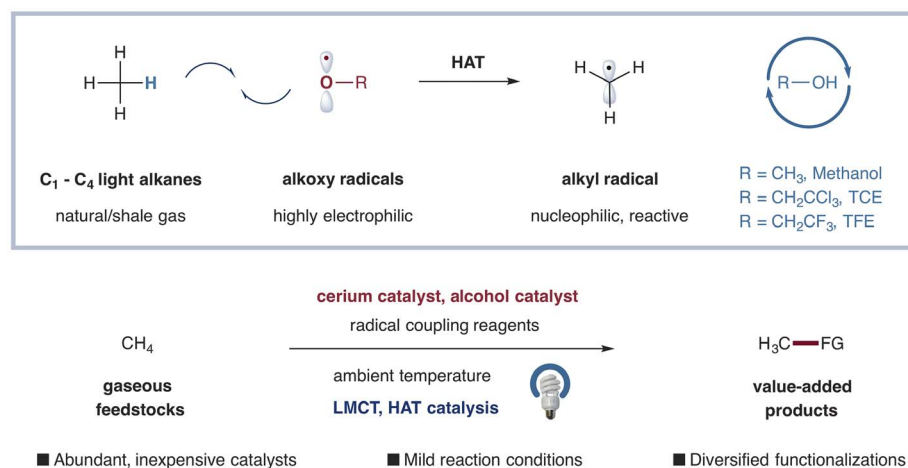
The intrinsic inertness of C–H bonds in methane and other gaseous alkanes has, however, brought extreme challenges for catalytic systems, not only in the activation step, but also in controlling chemoselectivity to avoid solvent functionalization and overfunctionalization under the frequently harsh conditions necessary (high temperature, superacid media, strong oxidants) (9). Moreover, the gaseous substrates' low solubility in most solvents has raised substantial practical difficulties. Elegant catalytic systems using transition metals such as Pd, Ir, Rh, and Ru have been reported recently (5, 10–12); however, the challenge remains to develop efficient Earth-abundant catalytic systems that operate under ambient conditions (1).

Photoredox catalysis has recently emerged as a powerful platform for the direct activation and functionalization of organic molecules via open-shell pathways (13). Hydrogen atom transfer (HAT) by photocatalytically generated amine and aminium (14–17), thiol (18), and other heteroatom-

centered radicals (19, 20) has recently enabled myriad otherwise unattainable C–H functionalization transformations under mild reaction conditions (21). The unique reactivity and polarity-governed selectivity (22) inherent in the abstraction events by these heteroatom-centered radicals inspired us to question whether a photocatalytic HAT strategy could be extended to the most challenging  $C(sp^3)$ –H transformations: the selective functionalization of methane and other gaseous alkanes. The nucleophilic character of alkyl radicals would provide strategic advantages for the development of diverse selective functionalizations, complementary to transition metal-catalyzed reactions, under mild conditions. Although gas-phase HAT events between methane and open-shell metal-oxo species have been extensively investigated for oxidative couplings of methane

at elevated temperature (23), room-temperature light alkane functionalization via HAT photocatalysis in a homogeneous system remains largely unresolved.

Ligand-to-metal charge transfer (LMCT) is a common photoexcitation manifold among coordination complexes of transition metals with an empty valence shell (24), which has nevertheless been underinvestigated and underutilized in synthetic organic transformations via modern photoredox catalysis (25, 26). We recently applied LMCT catalysis to the direct activation of alcohol feedstocks, enabling alkoxy radical-mediated skeletal rearrangement of cyclic alcohols (27) and remote C–H functionalization of primary alcohols via 1,5-HAT (28). Using absorbed light energy to promote targeted oxidation via the homolysis of the transiently coordinated  $Ce(IV)$ –alkoxide, challenging oxidations of a wide range of alcohols have been accomplished under mild and operationally simple conditions with cerium(III) salts as precatalyst. Recognizing that the ready abundance of cerium would make it an ideal catalyst for sustainable, large-scale photocatalytic systems (29, 30), we hypothesized that the synergistic merger of LMCT and HAT catalysis, via the use of inexpensive cerium salts in concert with simple alcohols such as methanol or 2,2,2-trichloroethanol, could productively activate and functionalize inert alkanes such as methane and ethane (Fig. 1). Indeed, highly electrophilic alkoxy radicals [ $CH_3O\cdot$ , O–H bond dissociation energy (BDE) = 105 kcal/mol] would be thermodynamically viable for the abstraction of C–H bonds from methane (C–H BDE = 105 kcal/mol) (31). Conceivably, the polarity-matching effect would render the hydrogen atom transfer with the ubiquitous, weak and acidic C–H bonds of solvent ( $CH_3CN$ , C–H BDE = 93 kcal/mol) kinetically disadvantageous, thereby subverting the frequently encountered problem of competitive solvent functionalization in methane functionalization. Critically, the use of inexpensive and widely available alcohols as HAT



**Fig. 1. Alkoxy radical-mediated HAT activation enables diverse selective functionalizations of gaseous alkanes.** The synergistic merger of LMCT and HAT catalysis would offer opportunities for the use of methane and other gaseous alkanes under economical ambient conditions. TCE, 2,2,2-trichloroethanol; TFE, 2,2,2-trifluoroethanol; FG, functional group.

School of Physical Science and Technology, ShanghaiTech University, Shanghai 201210, China.

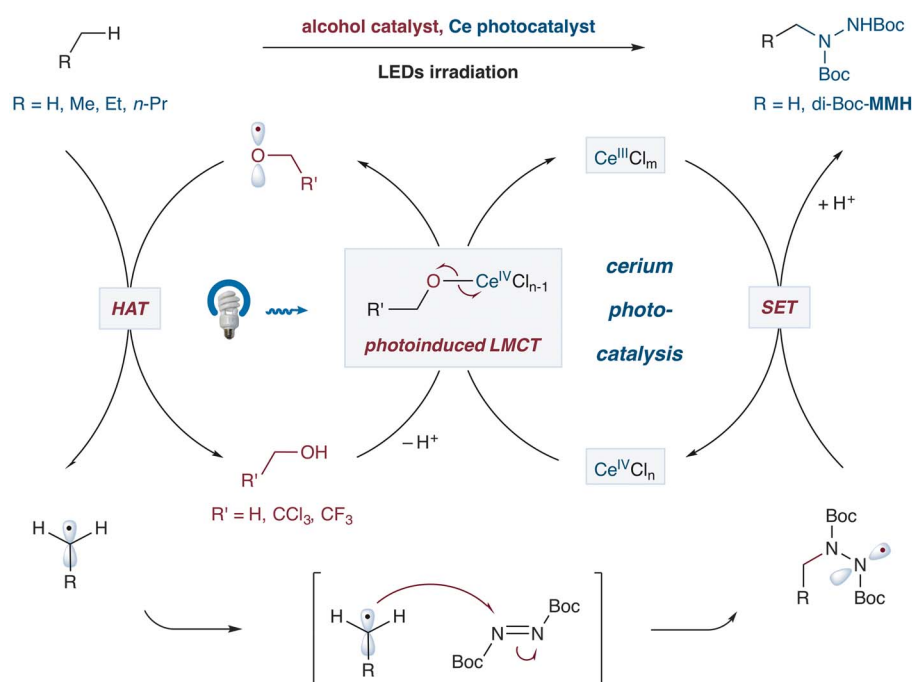
\*These authors contributed equally to this work.

†Corresponding author. Email: zuozhw@shanghaitech.edu.cn

catalysts would allow easy tunability of reactivity and selectivity (32). Here, we report the application of LMCT-enabled HAT catalysis as a general method for the catalytic and selective radical coupling of methane, ethane, and several other simple alkanes.

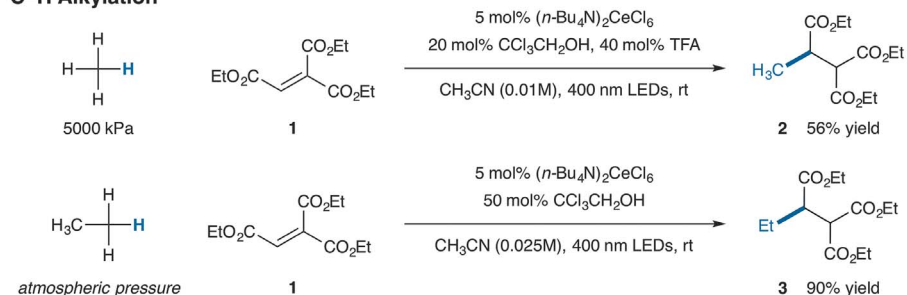
We sought first to apply our LMCT/HAT functionalization to the amination of methane, a largely elusive transformation (33). The direct transformation of methane into liquid *tert*-butoxycarbonyl (Boc)-protected monomethylhydrazine (MMH)—an important rocket propellant and valuable building block for heterocyclic compounds, currently produced in industrial quantities via oxidation of methylamine—would upgrade methane from an abundant feedstock into a useful commodity chemical. A detailed description of our proposed catalytic mechanism is outlined in Fig. 2. We envisioned that a Ce(IV)-alkoxy complex, catalytically generated in situ from simple alcohols and a Ce(IV) salt, would undergo photoinduced LMCT to generate a high-energy electrophilic alkoxy radical and a reduced Ce(III) species. The alkoxy radical would then abstract an H atom from a C–H bond in the substrate alkane to generate an alkyl radical species, which would then readily couple with di-*tert*-butyl azodicarboxylate (DBAD) to forge a new C–N bond with the formation of a stable captodative N-centered radical. Single electron reduction of this radical by the reduced form of the cerium catalyst would regenerate the active Ce(IV) and deliver the desired product after protonation.

In practice, the photocatalyzed reaction was carried out in a standard pressure reactor equipped with a sapphire window in the top plate to transmit 400-nm LED irradiation (see fig. S2). At ambient temperature under 5000 kPa of methane with DBAD as the limiting reagent, we observed that using 0.5 mole percent (mol %) CeCl<sub>3</sub>, in concert with catalytic 2,2,2-trichloroethanol, the desired Boc-protected MMH product was formed in 39% yield (table S6, entry 1; see tables S6 to S10 for detailed studies). Among several commercial cerium salts we evaluated in combination with exogenous soluble tetrabutylammonium chloride, cerium(IV) trifluoromethanesulfonate gave slightly higher efficiency (Table 1, entry 1); the premade Ce(IV) chloride complex (34) also proved effective in this catalytic system (table S6, entry 2). No overfunctionalization product was observed in any cases, and the product proved to be inert when subjected to the reaction conditions (fig. S7). Moreover, the alcohol HAT catalyst 2,2,2-trichloroethanol was left largely intact at the end of the reaction. The major unproductive side pathway was the reduction of DBAD to the corresponding hydrazine. An increase in absolute yield could be obtained by using D<sub>3</sub>-acetonitrile, affording di-Boc-MMH in 63% yield, as a smaller amount of reduced hydrazine by-product was formed (Table 1, entry 2). Moreover, no significant decrease in efficiency was observed when the reaction was performed at 0°C (table S9, entry 2), highlighting the highly reactive nature of the radical coupling system. Lowering the catalyst loading to 0.1 mol % or even 0.01 mol % increased the

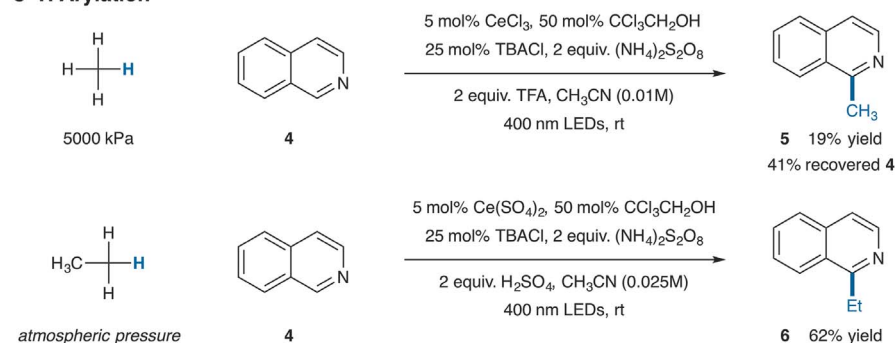


**Fig. 2. Proposed mechanism for the cerium-catalyzed C(sp<sup>3</sup>)-H functionalization of methane and other gaseous alkanes.** Highly electrophilic alkoxy radicals, generated from simple alcohols via photoinduced LMCT, are used for the HAT activation of challenging C–H bonds in light alkanes. di-Boc-MMH, di-*tert*-butoxycarbonyl monomethylhydrazine. Me, methyl; Et, ethyl; Pr, propyl.

### C–H Alkylation



### C–H Arylation



**Fig. 3. Photocatalytic alkylations using methane and ethane.**

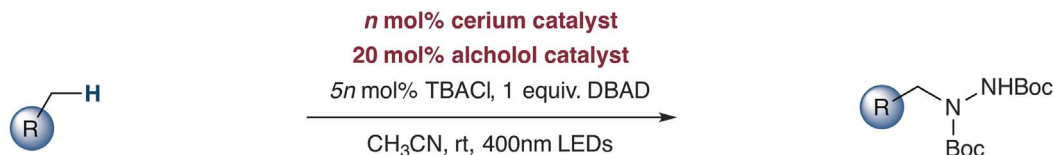
turnover numbers (TONs) to 320 and 2900, respectively, although this necessitated longer reaction times and resulted in minimally decreased yield (Table 1, entry 3, and table S8, entry 3). Control experiments indicated the necessity of cerium

catalyst, alcohol catalyst, and LED irradiation, as omission of any of these resulted in no desired amination product (table S1).

The catalyst combination was also highly effective for ethane functionalization (see tables S11



**Table 1. Cerium-catalyzed amination of alkanes.** Yields and regiomer ratios (r.r.) of all products were determined by gas chromatography analysis with internal standard. Regiomer ratios are presented as 1°/2° product ratios. See full details in supplementary materials. \*With 40 mol % trifluoroacetic acid. †In D<sub>3</sub>-acetoneitrile, in the absence of TBACl and trifluoroacetic acid. ‡With diisopropyl azodicarboxylate (DIAD). §At 1000 kPa of ethane. ||With 50 mol % alcohol catalyst. rt, room temperature; TBACl, tetrabutylammonium chloride; DBAD, di-*tert*-butyl azodicarboxylate.



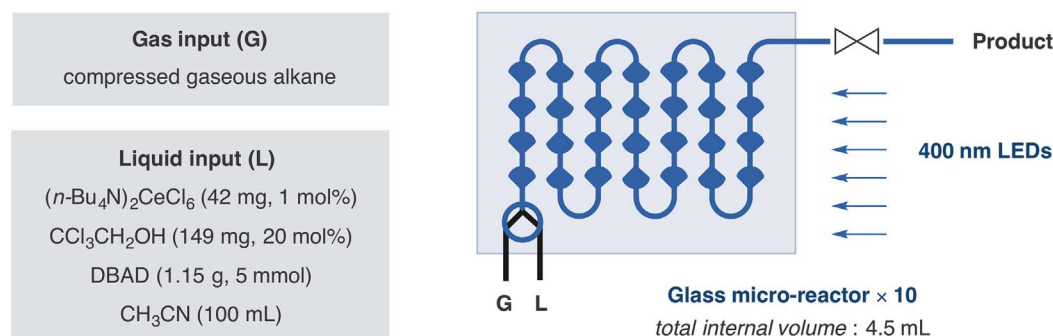
| Alkane                              | Entry | Cerium catalyst                                               | Loading (n) | Alcohol catalyst                     | Time | Yield            | TON         |
|-------------------------------------|-------|---------------------------------------------------------------|-------------|--------------------------------------|------|------------------|-------------|
| R = H*<br><b>Methane</b> (5000 kPa) | 1     | Ce(OTf) <sub>4</sub>                                          | 0.5         | CCl <sub>3</sub> CH <sub>2</sub> OH  | 2 h  | 45%              | <b>90</b>   |
|                                     | 2†    | ( <i>n</i> -Bu <sub>4</sub> N) <sub>2</sub> CeCl <sub>6</sub> | 0.5         | CCl <sub>3</sub> CH <sub>2</sub> OH  | 2 h  | 63%              | <b>126</b>  |
|                                     | 3     | Ce(OTf) <sub>4</sub>                                          | 0.01        | CCl <sub>3</sub> CH <sub>2</sub> OH  | 18 h | 29%              | <b>2900</b> |
| R = Me‡<br><b>Ethane</b> (101 kPa)  | 4     | CeCl <sub>3</sub>                                             | 0.5         | CCl <sub>3</sub> CH <sub>2</sub> OH  | 4 h  | 74%              | <b>148</b>  |
|                                     | 5§    | CeCl <sub>3</sub>                                             | 0.01        | CCl <sub>3</sub> CH <sub>2</sub> OH  | 4 h  | 97%              | <b>9700</b> |
| R = Et<br><b>Propane</b> (101 kPa)  | 6     | CeCl <sub>3</sub>                                             | 0.5         | CCl <sub>3</sub> CH <sub>2</sub> OH  | 9 h  | 70% (1:1 r.r.)   | <b>140</b>  |
|                                     | 7     | CeCl <sub>3</sub>                                             | 0.5         | CF <sub>3</sub> CH <sub>2</sub> OH   | 12 h | 61% (1:1 r.r.)   | <b>122</b>  |
|                                     | 8     | CeCl <sub>3</sub>                                             | 0.5         | CH <sub>3</sub> OH                   | 19 h | 39% (1:3.9 r.r.) | <b>78</b>   |
| R = Pr<br><b>Butane</b> (101 kPa)   | 9     | CeCl <sub>3</sub>                                             | 0.5         | CCl <sub>3</sub> CH <sub>2</sub> OH  | 6 h  | 76% (1:1.7 r.r.) | <b>152</b>  |
|                                     | 10    | CeCl <sub>3</sub>                                             | 0.5         | (CH <sub>3</sub> ) <sub>2</sub> CHOH | 18 h | 40% (1:4 r.r.)   | <b>80</b>   |
|                                     | 11    | CeCl <sub>3</sub>                                             | 0.5         | CH <sub>3</sub> OH                   | 18 h | 72% (1:8 r.r.)   | <b>144</b>  |
| <b>Cyclohexane</b>                  | 12    | CeCl <sub>3</sub>                                             | 0.5         | CCl <sub>3</sub> CH <sub>2</sub> OH  | 16 h | 81%              | <b>162</b>  |

and S12 for detailed studies). Because of ethane's higher solubility in acetonitrile relative to methane, the reactions could be performed in regular vials under atmospheric pressure. Satisfactory yield and TONs were obtained using 0.5 mol % CeCl<sub>3</sub> and 20 mol % 2,2,2-trichloroethanol (Table 1, entry 4). Furthermore, lowering the cerium catalyst loading from 0.05 mol % to 0.01 mol % led to catalytic TONs up to 6900 under ambient temperature and pressure conditions (table S12, entry 3). Additionally, increasing the pressure of ethane to 1000 kPa led to a remarkable rate acceleration, yielding 97% of aminated product in just 4 hours, at 0.01 mol % loading of cerium catalyst, corresponding to a TON of 9700 (Table 1, entry 5).

Propane and butane, the major components of liquid petroleum gas (LPG), a commonly used fuel, present interesting challenges with respect

to selectivity of C–H bond functionalization (35, 36). We hypothesized that the regioselectivity of the C–H bond abstraction step could be tuned by modification of the alcohol HAT catalyst (22, 32, 37, 38). In both propane and butane, the methylene positions possess slightly weaker C–H bonds, whereas the methyl positions possess more numerous C–H bonds. We were pleased to observe that with low loading of CeCl<sub>3</sub> and simple alcohols (2,2,2-trichloroethanol or 2,2,2-trifluoroethanol), propane could be functionalized to provide aminated products with high efficiency, although in a nonselective manner (Table 1, entries 6 and 7). Using methanol as the HAT catalyst, the regioselectivity improved to 1:3.9 (Table 1, entry 8), favoring the methylene position, although the reaction proceeded at a much lower rate. This selectivity could arise from the electronic difference in the electrophilic nature of methoxy

radical relative to halogenated ethoxy radicals. In a similar vein, butane could be functionalized with high efficiency, and the regioselectivity could be tuned with different choice of alcohol catalysts to favor the methylene positions. When using isopropanol as the HAT catalyst, the 1°/2° product ratio could be increased to 1:4 (Table 1, entry 10); the lower efficiency was likely due to the decomposition of the isopropanol catalyst via alkoxy radical-mediated β-scission, as the side product of methyl radical coupling with DBAD was observed in the reaction (table S14, entry 5). Methanol proved to be a selective HAT catalyst, providing 1:8 selectivity for the methylene C–H bonds (Table 1, entry 11); this result validated the hypothesis that enhancement of regioselectivity could be accomplished through electronic tuning of the alcohol HAT catalyst. Pleasingly, liquid hydrocarbons such as cyclohexane could be aminated



**Fig. 4. Photo-catalytic amination of alkanes in continuous-flow reactors for scaled-up applications.** \*At 0.01 M with 40 mol % trifluoroacetic acid. †With DIAD. ‡In liquid mode at 0.2 M.

| Entry | Alkane              | Flow rate   | Residence time | Yield | Productivity |
|-------|---------------------|-------------|----------------|-------|--------------|
| 1*    | methane (1800 kPa)  | 0.3 mL/min  | 15 min         | 15%   | 0.6 mmol/d   |
| 2†    | ethane (1500 kPa)   | 0.75 mL/min | 6 min          | 90%   | 2.0 mmol/h   |
| 3     | propane (800 kPa)   | 0.75 mL/min | 6 min          | 76%   | 1.7 mmol/h   |
| 4     | butane (400 kPa)    | 0.75 mL/min | 6 min          | 56%   | 1.3 mmol/h   |
| 5‡    | cyclohexane (22 mL) | 0.5 mL/min  | 9 min          | 70%   | 4.2 mmol/h   |

with high efficiency using this inexpensive photocatalytic platform at a low cerium loading (Table 1, entry 12).

We next applied this photocatalytic protocol to radical alkylation reactions using methane and ethane as alkylating feedstocks. As shown in Fig. 3, the alkoxy-mediated methylation and ethylation of electron-deficient alkene **1** using methane and ethane delivered the desired product with high efficiency (56% and 90% yield, respectively). Furthermore, methane and ethane proved viable pronucleophiles in Minisci arene alkylation reactions. The acidic conditions made use of diminished product overfunctionalization by electronically deactivating the benzylic C–H bonds through protonation of the N-heterocycle (39). With ammonium persulfate as an economical oxidant, methylated and ethylated isouquinoline could be produced at room temperature.

Photocatalytic reactions in batch reactors are typically limited to small-scale applications because of the attenuation effect of superficial light penetration (40). In addition to the increase of the light utilization efficiency through glass micro-reactors, the development of continuous-flow photocatalytic systems for large volume reactions could further benefit from the ease of handling gaseous reactants and the enhanced mass transfer between gas and liquid phases, which are crucial for the utilization of gaseous alkanes. Given the attractiveness of flow chemistry for mixed-phase gas/solution reactions, we investigated the C–H functionalization of gaseous alkanes in continuous-flow microreactors (Fig. 4). An acetonitrile solution of reagents and catalysts was pumped as a single stream and then directly combined and mixed with the gaseous alkane feedstock in the microreactor. Using a flow setup of 10 parallel microreactors (4.5 ml internal volume in total), at a pressure of 1500 kPa for

the ethane gas and a flow rate of 0.75 ml/min for the liquid solution stream, amination product was furnished in 90% yield with a residence time of 6 min and a production throughput of 2 mmol/hour. Under similar conditions, propane and butane were functionalized with good yield (76% and 56%, respectively) and productivity (1.7 mmol/hour and 1.3 mmol/hour). In our attempts to use methane in this mixed liquid/gas flow reaction setting, we were hampered by the pressure limits of our commercial microreactors (1800 kPa), leading to a somewhat diminished, but promising, 15% yield. As for higher alkanes such as cyclohexane, the scaled-up amination could be conveniently performed with simple liquid injection mode, affording remarkable productivity of 4.2 mmol/hour.

Further studies were performed to provide additional evidence for the intermediacy of alkoxy radicals under this catalytic manifold. As shown in Fig. S11, methoxy, trichloroethoxy, and trifluoroethoxy radicals could each be trapped by styrene to generate 1,2-alkoxyamination products (80%, 25%, and 59% yield, respectively). In the case of 2,2,2-trifluoroethanol, a small amount of trifluoromethylated product was also observed, indicative of a β-scission pathway and trapping of trifluoromethyl radical.

This photocatalytic platform has enabled several direct transformations of methane and other simple hydrocarbons, including amination, alkylation, and arylation, and offers intriguing opportunities for further functionalizations of feedstock alkanes.

#### REFERENCES AND NOTES

- N. J. Gunsalus *et al.*, *Chem. Rev.* **117**, 8521–8573 (2017).
- U.S. Energy Information Administration, *International Energy Outlook 2017*; www.eia.gov/outlooks/ieo/pdf/0484 (2017).pdf.
- A. E. Shilov, G. B. Shul'pin, *Chem. Rev.* **97**, 2879–2932 (1997).

- R. A. Periana *et al.*, *Science* **280**, 560–564 (1998).
- R. A. Periana, O. Mironov, D. Taube, G. Bhalla, C. J. Jones, *Science* **301**, 814–818 (2003).
- A. D. Sadow, T. D. Tilley, *J. Am. Chem. Soc.* **127**, 643–656 (2005).
- M. V. Kirilova *et al.*, *J. Am. Chem. Soc.* **129**, 10531–10545 (2007).
- A. Caballero *et al.*, *Science* **332**, 835–838 (2011).
- R. H. Crabtree, *Chem. Rev.* **95**, 987–1007 (1995).
- V. N. Cavaliere, B. F. Wicker, D. J. Mindiola, *Adv. Organomet. Chem.* **60**, 1–47 (2012).
- A. K. Cook, S. D. Schimler, A. J. Matzger, M. S. Sanford, *Science* **351**, 1421–1424 (2016).
- K. T. Smith *et al.*, *Science* **351**, 1424–1427 (2016).
- C. R. J. Stephenson, T. P. Yoon, D. W. C. MacMillan, *Visible Light Photocatalysis in Organic Chemistry* (Wiley, 2018).
- J. L. Jeffrey, J. A. Terrett, D. W. C. MacMillan, *Science* **349**, 1532–1536 (2015).
- G. J. Choi, Q. Zhu, D. C. Miller, C. J. Gu, R. R. Knowles, *Nature* **539**, 268–271 (2016).
- J. C. K. Chu, T. Rovis, *Nature* **539**, 272–275 (2016).
- C. Le, Y. Liang, R. W. Evans, X. Li, D. W. C. MacMillan, *Nature* **547**, 79–83 (2017).
- J. D. Cuthbertson, D. W. C. MacMillan, *Nature* **519**, 74–77 (2015).
- S. Mukherjee, B. Maji, A. Tlahuext-Aca, F. Glorius, *J. Am. Chem. Soc.* **138**, 16200–16203 (2016).
- K. A. Margrey, W. L. Czaplinski, D. A. Nicewicz, E. J. Alexanian, *J. Am. Chem. Soc.* **140**, 4213–4217 (2018).
- J. C. K. Chu, T. Rovis, *Angew. Chem. Int. Ed.* **57**, 62–101 (2018).
- B. P. Roberts, *Chem. Soc. Rev.* **28**, 25–35 (1999).
- N. Dietl, M. Schlagen, H. Schwarz, *Angew. Chem. Int. Ed.* **51**, 5544–5555 (2012).
- V. Balzani, P. Ceroni, A. Juris, *Photochemistry and Photophysics: Concepts, Research, Applications* (Wiley, 2014).
- S. J. Hwang *et al.*, *J. Am. Chem. Soc.* **137**, 6472–6475 (2015).
- B. J. Shields, A. G. Doyle, *J. Am. Chem. Soc.* **138**, 12719–12722 (2016).
- J. J. Guo *et al.*, *Angew. Chem. Int. Ed.* **55**, 15319–15322 (2016).
- A. Hu *et al.*, *J. Am. Chem. Soc.* **140**, 1612–1616 (2018).
- H. Yin, P. J. Carroll, J. M. Anna, E. J. Schelter, *J. Am. Chem. Soc.* **137**, 9234–9237 (2015).
- H. Yin *et al.*, *J. Am. Chem. Soc.* **138**, 16266–16273 (2016).
- S. J. Blanksby, G. B. Ellison, *Acc. Chem. Res.* **36**, 255–263 (2003).
- M. Salamone, M. Bietti, *Acc. Chem. Res.* **48**, 2895–2903 (2015).



33. D. Michos, C. A. Sassano, P. Krajnik, R. H. Crabtree, *Angew. Chem. Int. Ed. Engl.* **32**, 1491–1492 (1993).
34. J. L. Ryan, C. K. Jørgensen, *J. Phys. Chem.* **70**, 2845–2857 (1966).
35. G. C. Fortman *et al.*, *J. Am. Chem. Soc.* **136**, 8393–8401 (2014).
36. B. G. Hashiguchi *et al.*, *Science* **343**, 1232–1237 (2014).
37. V. A. Schmidt, R. K. Quinn, A. T. Brusoe, E. J. Alexanian, *J. Am. Chem. Soc.* **136**, 14389–14392 (2014).
38. R. K. Quinn *et al.*, *J. Am. Chem. Soc.* **138**, 696–702 (2016).
39. M. Salamone, G. Carboni, M. Bietti, *J. Org. Chem.* **81**, 9269–9278 (2016).
40. D. Cambié, C. Bottecchia, N. J. Straathof, V. Hessel, T. Noël, *Chem. Rev.* **116**, 10276–10341 (2016).

#### ACKNOWLEDGMENTS

**Funding:** We thank the National Natural Science Foundation of China (21772121) and the “Thousand Plan” Youth program for financial support. **Author contributions:** Z.Z. conceived and directed the project; Z.Z., A.H., J.J.G., and H.P. designed the experiments; A.H., J.J.G., and H.P. performed and analyzed the photocatalytic batch reactions; H.P. performed continuous-flow reactions; and Z.Z., A.H., J.J.G., and H.P. prepared the manuscript.

**Competing interests:** The authors declare no conflicts of

interest. **Data and materials availability:** Data are available in the supplementary materials.

#### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/361/6403/668/suppl/DC1](http://www.sciencemag.org/content/361/6403/668/suppl/DC1)  
Materials and Methods  
Figs. S1 to S11  
Tables S1 to S15  
GC data and NMR spectra  
2 May 2018; accepted 3 July 2018  
Published online 26 July 2018  
10.1126/science.aat9750

## ACTIVE MATTER

# Collective clog control: Optimizing traffic flow in confined biological and robophysical excavation

J. Aguilar<sup>1\*</sup>, D. Monaenkova<sup>2\*</sup>, V. Linevich<sup>1</sup>, W. Savoie<sup>2</sup>, B. Dutta<sup>3</sup>, H.-S. Kuan<sup>4</sup>, M. D. Betterton<sup>5</sup>, M. A. D. Goodisman<sup>6</sup>, D. I. Goldman<sup>2†</sup>

Groups of interacting active particles, insects, or humans can form clusters that hinder the goals of the collective; therefore, development of robust strategies for control of such clogs is essential, particularly in confined environments. Our biological and robophysical excavation experiments, supported by computational and theoretical models, reveal that digging performance can be robustly optimized within the constraints of narrow tunnels by individual idleness and retreating. Tools from the study of dense particulate ensembles elucidate how idleness reduces the frequency of flow-stopping clogs and how selective retreating reduces cluster dissolution time for the rare clusters that still occur. Our results point to strategies by which dense active matter and swarms can become task capable without sophisticated sensing, planning, and global control of the collective.

**D**iverse living (1) and artificial (2) active materials (3) and swarms spontaneously form clusters that can persist for long durations. However, for tasks that demand steady flow, such formations can be disadvantageous: Confined active systems such as pedestrian or vehicular traffic jams (4), competing bacterial biofilms (5), high-density migrating cells (6), jammed herds (7), and robot swarms (8) can produce high-density clogs that readily form glasslike arrests of flow (9). In such systems, the ability to dissolve clusters and prevent their formation (9), particularly in the absence of global knowledge of the state of all elements, is crucial.

Social insects (10) perform many tasks that demand clog minimization and mitigation. Substrate excavation specialists such as fire ants (*Solenopsis invicta*) cooperatively create nests of complex subterranean networks (Fig. 1A) consisting of tunnels in soil that support bidirectional traffic without lanes (11). Our previous laboratory experiments (12) revealed that, in the early stages of nest construction, the few-millimeter-long ants construct vertical tunnels approximately one body length in diameter (13). These narrow tunnels benefit the climbing ants as they transport bulky pellets, because close proximity to walls allows limbs, body parts, and antennae to aid slip recovery (12). But although

the structure of the tunnels seems to benefit individuals, physical-model experiments make it clear that excavation can suffer as a result of clogging during high-traffic conditions [e.g., (14) and Fig. 1C]. Here we use biological, theoretical, computational, and robophysical systems to show that counterintuitive behaviors—individual idleness and retreating—help optimize tunnel density by limiting the severity and prevalence of clogs, thereby enabling rapid excavation by the collective.

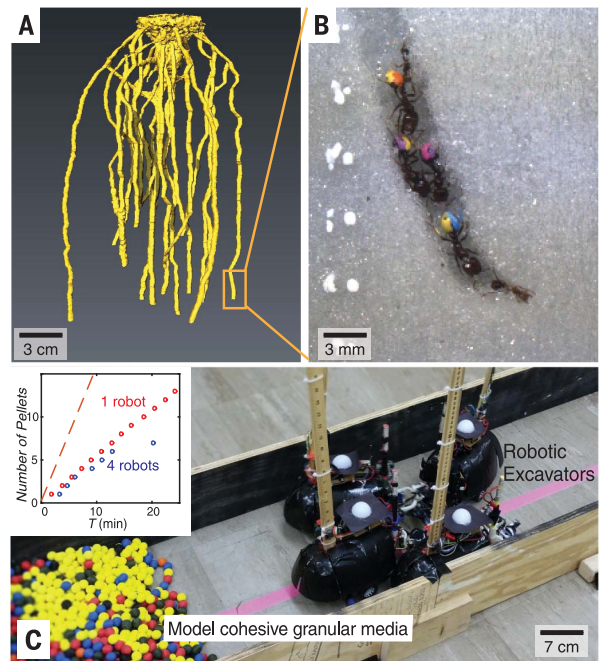
In laboratory experiments, we monitored the activity of fire ants as they excavated a cohesive

granular medium. Groups of ~30 workers were placed in transparent containers containing particle-water mixtures (13) consisting of 0.25-mm-diameter glass particles (Fig. 2A) with a soil moisture content, defined as the ratio of total water weight to total solid weight, of 0.01 or 0.1 (three trials each) (13). Ants excavated for 48 hours, with individual ants entering and exiting the tunnel hundreds of times. As in our previous study (12), ants constructed narrow vertical tunnels by means of a stereotyped process of grain and multigrain (pellet) removal and transport, followed by tunnel ascent and substrate deposition upon exit (13). A camera mounted to a motorized linear stage tracked a region within about three body lengths from the tunnel face (Fig. 2A and supplementary materials). We distinguished individual ant activity by marking ant abdomens with different colors (Fig. 1B). We recorded tunnel length over time (Fig. 2B), and the presence of each worker was logged when in the camera's view (Fig. 2C).

Ants exhibited a variety of behavioral tasks during collective excavation. A large fraction ( $0.22 \pm 0.1$  for soil moisture content of 0.01 and  $0.31 \pm 0.13$  for soil moisture content of 0.1) of ants never entered the tunnel to excavate during the 48-hour period of observation; we refer to these as “nonvisitors.” As seen in Fig. 2C, ants that visited the tunnel face (“visiting” ants) varied in activity level. Inspired by work in honey bees (15), we quantified activity inequality among visitor ants using Lorenz curves. Points on the Lorenz curves in Fig. 2D link the cumulative fraction of workers in the population to the cumulative share of activity by that fraction. Although visitor ants' trips did not always result in the extraction of a pellet (see movie S1 and discussion

**Fig. 1. Confined and crowded biological and robotic excavators.**

(A) X-ray reconstruction of *S. invicta* fire ant excavation in a large container (25 cm wide) filled with 240- to 270- $\mu$ m-diameter glass particles (supplementary materials). (B) Painted *S. invicta* workers excavating a single tunnel along the wall of a transparent container with 0.25-mm-diameter wet glass particles. (C) Autonomous robotic diggers excavating in a simulated environment with cohesive granular media (diameter of 1.8 cm). The inset shows the number of pellets (defined as a cohesive group of grains) deposited versus time (7) by a robot excavating alone (red dots) and the net excavation of four robots (blue circles), whereby each robot attempts to excavate maximally. Orange dashed line indicates the hypothetical performance of the group of four robots in the absence of confinement.



<sup>1</sup>School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA. <sup>2</sup>School of Physics, Georgia Institute of Technology, Atlanta, GA 30332, USA.

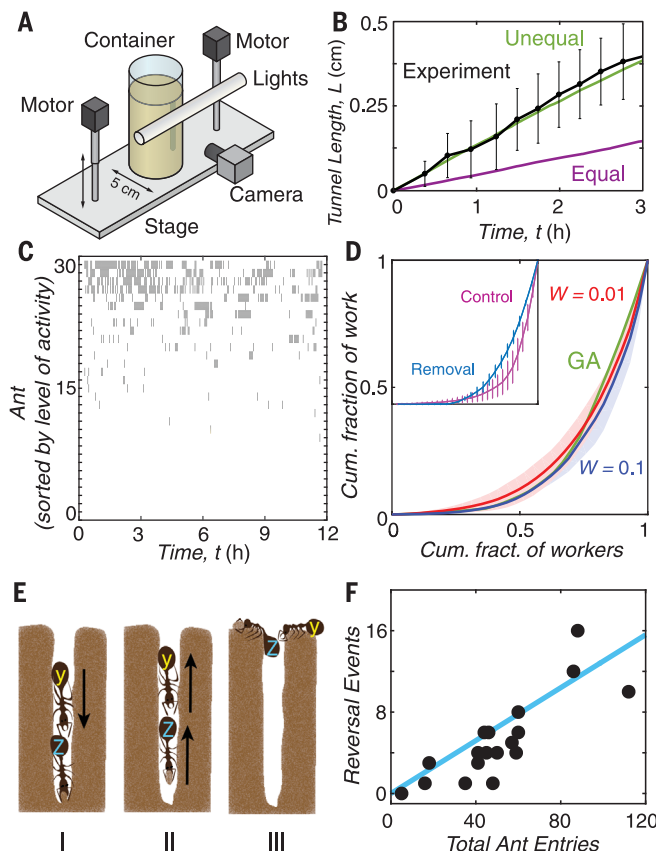
<sup>3</sup>School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA. <sup>4</sup>Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Str. 38, 01187 Dresden, Germany. <sup>5</sup>Department of Physics, University of Colorado Boulder, Boulder, CO 80309, USA. <sup>6</sup>School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA.

\*These authors contributed equally to this work.  
†Corresponding author. Email: daniel.goldman@physics.gatech.edu

## Fig. 2. Biological observations reveal workload inequality and reversal behaviors in ants.

(A) Experimental apparatus to track ant excavation; the inner diameter of the container is 5.21 cm. (B) The growth of tunnel length over time. Shown are average experimental results  $\pm$  SD/2 for *S. invicta* workers (black) and simulations for groups with equal (purple) and unequal (green) workload distribution. Error bars denote 1 SD in each direction. (C) “Visitation” map derived from experimental data. Each point in the map indicates the presence of a particular ant (out of 30 ants), ordered from most active (y axis) in the tunnel at a time  $t$  [soil moisture content ( $W$ ) of 0.1]. (D) Lorenz curves for workload distributions obtained in wet

0.25-mm-diameter glass particles with soil moisture content of 0.1 (blue) and 0.01 (red) and a CA model (green) whose excavation rate was optimized with a GA. Shaded areas correspond to standard deviation from three experiments. cum., cumulative. The inset shows average Lorenz curves  $\pm$  SD/2 for a workload distribution within the group before (control, purple) and after (removal, blue) the most active diggers are removed from the group. Error bars correspond to standard deviations from three experiments. (E) Illustration of observed reversal behavior. (I) Ant Y’s path to excavate is blocked by ant Z. (II) After Z collects a pellet, it reverses, (III) forcing Y to reverse without excavating. (F) Total number of reversal events versus total ant visitors for the first 3 hours of ant excavation (soil moisture content of 0.1). Each data point represents total reversal events and total entries counted for 30-min segments collected from three experiments. Linear fit (blue line) with coefficient of determination ( $R^2$ ) = 0.69.



below), we included these “reversals” in the Lorenz curve calculations because these animals expended energy in a trip to the tunnel face and contributed to tunnel traffic.

To characterize the Lorenz distributions, we calculated the Gini coefficient,  $G$ , defined as the ratio of the area between the Lorenz curve and the line of equality to the area under the line of equality (15).  $G$  is a measure of the deviation of the workload from perfectly shared ( $G = 0$ , all workers work equally) to completely unshared ( $G = 1$ , a single worker performs all work). Lorenz curves were characterized by  $G = 0.75 \pm 0.10$  and displayed similar functional forms across a variety of experimental conditions (see Fig. 2D and fig. S2).

In the presence of competing tasks, like foraging or brood care, task allocation in ants can change depending on colony needs (16). To investigate temporal variation in ant excavation workload, we divided 48-hour experiments into

12-hour “epochs” (time periods). Although individual activity varied among epochs (Fig. 2C and fig. S3), the cumulative workload distribution was independent of epoch [one-way analysis of variance (ANOVA)  $F_{3,20} = 0.85$ ,  $P = 0.48$ ] and soil moisture content (one-way ANOVA,  $F_{1,23} = 2.54$ ,  $P = 0.13$ ) (Fig. 2D, figs. S1 and S2, and table S1). Furthermore, when the most active excavators were removed from the group, remaining workers increased their activity and compensated for the loss, preserving the shape of the Lorenz curve and therefore producing similar Gini coefficients (one-way ANOVA,  $F_{1,4} = 1.13$ ,  $P = 0.35$ ) (Fig. 2D, inset; table S2; and supplementary materials). Thus, given the consistency of the workload distribution, we hypothesize that variations in idleness (low activity levels) within a population may play an adaptive role in modulating the crowded conditions of confined tunnels and could have been important in the earliest social insect colonies (17).

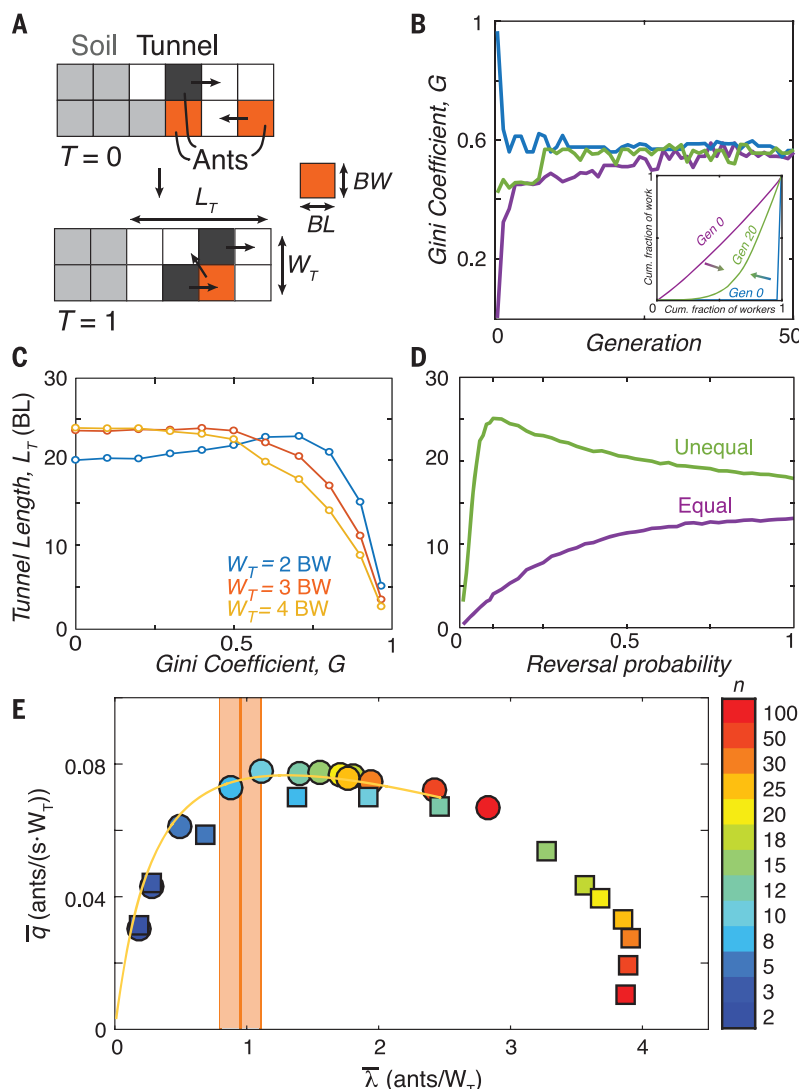
Reversal behaviors were characterized by ants entering the tunnel and returning to the exit without carrying soil pellets. During the first 3 hours of the experiments, reversals occurred for  $26 \pm 13\%$  of trips for soil moisture content of 0.01 and  $18 \pm 3\%$  of trips for soil moisture content of 0.1. These events were often associated with local crowding at the excavation face (Fig. 2E) ( $16 \pm 12\%$  of trips for soil moisture content of 0.01 and  $10 \pm 2\%$  of observations for soil moisture content of 0.1). Reversal behaviors in crowded conditions occur on foraging trails (18), and similar phenomena have been observed in swarming bacteria (5). The incidence of this seemingly unproductive behavior increased with increasing overall activity of ants (Fig. 2F), suggesting that this behavior serves as a feedback mechanism for mitigating clogs during excavation.

To systematically examine the effects of idleness and individual retreating behaviors on excavation performance, we developed a cellular automata (CA) excavation model (Fig. 3A and supplementary materials). Such models are useful in elucidating the dynamics of biological and vehicular traffic (9, 19). The model consists of a lattice (the “tunnel”) with a width of two cells [similar to *S. invicta* tunnel widths (20)] occupied by soil, empty space, an ascending CA “ant,” and/or a descending CA ant (Fig. 3A). The CA ants can move, change directions, excavate, deposit a pellet, or rest. As in the biological experiments, activity for the workload distribution in the CA model was measured by counting instances when CA ants visited the tunnel within three body lengths (cells) of the excavation site.

We simulated the behavior of CA ants using both equal workload distributions (which we refer to as “active” CA ants) and unequal workload distributions (which we refer to as “Lorenz” CA ants) with identical reversal probabilities (movie S2). In unequal workload distributions, individual CA ants were assigned individual “entrance probabilities” defined as the probability that a CA ant will enter the tunnel. The initial entrance probability distribution for the 30 CA ants was taken from the biological distribution. Output workload distributions of CA simulations closely matched the input entrance probability distributions (as measured by the Gini coefficient, fig. S25). During a time-step, if its path toward the excavation area was blocked, a CA ant would reverse direction toward the exit with a probability,  $R$ , of 0.34 (supplementary materials);  $R$  was set by the proportion of total reversal events observed for 0.01 soil moisture in the biological experiments.

The CA model that used unequal workload distribution and reversals reproduced experimentally observed biological ant digging rates (Fig. 2B). To determine if these rates represented an optimal workload distribution, we used a genetic algorithm (GA) (fig. S24) to select for entrance probability distributions (supplementary materials) that maximized excavated tunnel length within a given duration. Regardless of the initial population distribution (either similar to the ants or highly unequal), within a





**Fig. 3. Models reveal optimized traffic flow in narrow tunnels by means of selective retreating and workload inequality.** (A) Schematic showing the main components of the CA model. Cell colors denote soil (light gray), tunnel (white), ants moving toward the excavation site (orange), and ants exiting the tunnel (dark gray).  $T$ , simulation time-step. (B) Gini coefficient over time under GA optimization for groups started with a completely equal (purple), completely unequal (blue), and random (green) workload distribution. Lorenz curves (inset) for groups that begin with complete equality or inequality rapidly reach a similar workload distribution. (C) Excavated tunnel length,  $L_T$ , after 24-hour simulation time versus Gini coefficient for tunnels of different widths,  $W_T$ , for a 30-CA ant population. BL, body length; BW, body width. (D) Excavated tunnel length after 24-hour simulation time versus reversal probability for equal and unequal (optimized for 30 CA ants) workload distributions. (E) Simulated traffic flow ( $\bar{q}$ , number of ants divided by time in seconds times tunnel width) versus CA ant occupancy ( $\bar{\lambda}$ , number of ants divided by tunnel width, measured in excavator body widths) for groups of equally (squares) and unequally (circles) active ants. Color bar indicates the size,  $n$ , of the excavating group. The theoretical fundamental diagram of the OAT model (yellow curve) illustrates the need to limit tunnel traffic to one worker per body width of tunnel width to optimize flow and prevent deleterious clogs. Experimental ant observations reveal an average occurrence around this density (orange-shaded region, where the orange centerline is the mean and the extents are one standard deviation away from the mean).

few generations, the GA simulation converged to an unequal workload distribution (Fig. 3B, for a 30-ant example), which was similar to the experimentally observed biological workload distributions (Fig. 2D, green).

The CA model also revealed the importance of the reversal behavior in conjunction with unequal workload distributions. Although the active excavation could be improved by sufficient reversal probability, only a small amount of reversal was

needed to increase the excavation performance in the unequal distribution (Fig. 3D). Thus, in addition to the benefits narrow tunnels provide for climbing and pellet transport (12, 13), we hypothesize that the ants benefit from narrow tunnels by expending less energy to dig wider tunnels to the same depth. Such benefits would be useful in the early stages of new nest construction (e.g., after the colony is flooded out) during which establishing the colony underground is critical.

To gain insight into other benefits and constraints set by such narrow tunnels, we simulated 30 CA ants with varied workload distributions (characterized by distinct Gini coefficients) in tunnels of different widths. These distributions were created through a randomized Monte Carlo process, such that the Lorenz curves resulted in desired Gini coefficients. A peak in excavated length,  $L$ , versus Gini coefficient was observed in a tunnel two cells wide (Fig. 3C). Wider tunnels (three and four cells wide) resulted in broader performance peaks, indicating a decreased sensitivity in performance owing to workload distribution. This indicates that use of a narrow tunnel necessitates the “discovery” of the unequal workload distribution of ants.

We hypothesized that the unequal workload distribution and reversals were linked to uniform flow of CA ants in the tunnel. We therefore measured the average flow rate of successful excavators,  $\bar{q}$ , versus the average tunnel-width-normalized occupancy of excavators,  $\bar{\lambda}$  (the ratio of average number of ants in the tunnel to tunnel width measured in ant body widths). To generate a wide range of average occupancies, we varied the population size of the CA system.

The flow rate was optimal at an intermediate occupancy (Fig. 3E). This nonmonotonic trend in  $\bar{q}$  versus  $\bar{\lambda}$  is characteristic of various multi-agent systems, including bridge-building army ants (21) and vehicle traffic (22, 23), and is referred to in traffic literature as the “fundamental diagram” (24). Active ants, which do not modulate their workload distribution, increase tunnel occupancy with increasing population and thus exhibit optimal flow rates for only a few population sizes. By contrast, GA-optimized Lorenz ants produced tunnel occupancies in the ideal range by generating increasingly unequal workload distributions for increasing CA ant population sizes. Of particular importance, fire ants produced tunnel densities in the ideal range (Fig. 3E, orange-shaded region).

The ability of the ants to operate at the optimum in the fundamental diagram and the rapidity by which the GA model converges (Fig. 3B) indicate the existence of a simple governing principle for traffic control in confined task-oriented systems. To elucidate this principle, we formulated a minimal model of ant traffic in the narrowest (single-lane) tunnel: the one-at-a-time (OAT) model. This model, which builds on recent work on traffic of motor proteins on microtubules (25), allows us to estimate analytically how the excavation rate varies with the rate of ants entering the tunnel (supplementary materials) for various work-distribution strategies.

In the OAT model, ants enter the tunnel and move toward its face; descending ants reverse direction if they either reach the end of the tunnel or collide with an ant moving in the other direction (supplementary materials). We initially modulated occupancy by varying the entrance probability of all ants equally; as in the CA model, the flow rate of the OAT model was optimal at an intermediate width-normalized ant occupancy (Fig. 3E, yellow curve, and supplementary materials)—in particular, one excavator for every excavator that can fit along the width of the tunnel. Although the peak in the fundamental diagram has been associated with the transition between steady flow and propagating traffic jams (24), the OAT model highlights a key feature of confined tunnel excavation: Traffic dynamics are driven by tunnel width. Given the task-oriented nature of the system, successful traffic flow is only possible if a worker can travel the entire length of the tunnel and back. Thus, if there are enough workers in the tunnel to clog the path to or from the excavation site, traffic is likely to slow down. The OAT model highlights this scenario, as such clogs are unavoidable if more than one ant is in the single-lane tunnel. Because ants cannot pass each other or change lanes, only the first ant to enter can reach the end to excavate, whereas other workers collide with the first worker, reverse, and impede traffic.

Mechanisms that target a specific number of excavators occupying the tunnel given the tunnel's width promote ideal traffic flow. When individual ants in the OAT model were programmed to modulate their rate of reentry according to how often they reversed without excavating, the OAT model rapidly converged to Lorenz curves similar to the biological and GA-optimized CA ants (supplementary materials and fig. S23). Such rapid convergence highlights the benefit of targeting a specific number of ants (in this case, by establishing unequal workload distributions) in narrow tunnels.

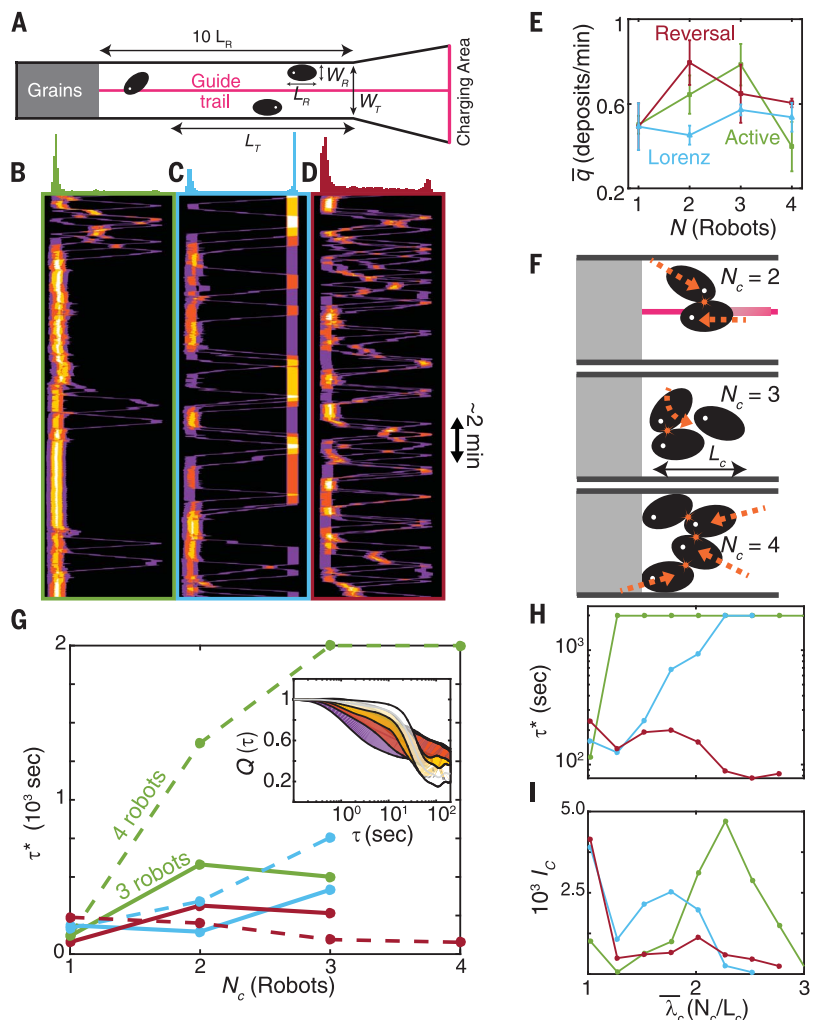
We next used a system of excavating robots (fig. S9) to test if the above theoretical strategies could improve traffic in confined experimental situations with more complex, unpredictable interactions. Because, presently, robot mobility in real-world environments is poor relative to biological systems and because real collisional interactions not modeled in CA and OAT are typically neglected in swarming robot studies (2), such robophysical (26) studies can aid robot design and control for real-world robot swarms, as well as suggest hypotheses for studies of ant traffic (18), adaptive behaviors, and morphological features for crowded excavation and movement.

Groups of roughly elliptical robots (movie S3) with similar aspect ratios to the biological ants were tasked with excavating a model cohesive granular medium of hollow plastic spheres containing loose magnets; this design allows clumps of media to be formed, analogous to the pellets of cohesive soil formed by the biological ants (13). Our robots followed simple instructions

triggered by onboard sensory feedback of the surrounding environment (supplementary materials). Previous work in swarm robotics (27) used similar decentralized strategies in conjunction with collision-avoidance schemes (2, 28) to produce emergent flocking behavior. By con-

trast, our robots detected collisions with push switches on their outer shell, which triggered navigation strategies such as steering away and readjusting to promote clog resolution (movie S4).

To challenge the robots, we constructed a tunnel (Fig. 4A) with a width of three robot



**Fig. 4. Traffic flow and local dynamics during robot excavation.** (A) Schematic of the excavation arena indicating the tunnel length,  $L_T$  (excluding the excavation area); robot width,  $W_R$ ; robot length,  $L_R$ ; and tunnel width,  $W_T$ . A pink centerline along the tunnel was monitored by the robots' onboard cameras, enabling them to follow the tunnel path. (B to D) Experimental space-time overlap heat maps of robot positions (x axis) for four-robot trials of (B) active digging, (C) Lorenz digging, and (D) reversal digging. Color indicates the number of robots occupying a particular space and time: one (purple), two (orange), three (yellow), and four (white) robots. Histograms above the graphs show the frequency of occurrence of clusters with two or more robots at different lateral positions. (E) Average flow rate,  $\bar{q}$ ,  $\pm$  SD measured in deposits per minute versus number of robots in the experiment,  $N$ , for active (green), Lorenz (light blue), and reversal (maroon) strategies. (F) Illustration of various collision scenarios encountered by robots owing to movement toward guide trail (top), turning (middle), and forward-backward translation (bottom). Orange starbursts indicate collisions. (G to I) Relaxation times for all strategies: active (green), Lorenz (light blue), and reversal (maroon). (G) Relaxation time versus cluster size,  $N_c$ , for three-robot (solid) and four-robot (dashed) trials. The inset shows sample average correlation curves,  $Q(\tau)$ , that measure how  $N_c = 1$  (purple), 2 (orange), 3 (yellow), and 4 (white) robot clusters dissolve over time during four-robot reversal trials; shaded region indicates average curves  $\pm$  SD/2 (standard deviations for  $\tau^*$  range from 100 to 500 s). (H) Relaxation times versus linear aggregation density,  $\bar{\lambda}_c$ , for four-robot trials and (I) corresponding number of cluster occurrences,  $I_c$ , versus linear aggregation density.

widths (or 1.5 robot lengths), which, combined with the oblong robot shape, forced a challenge of turning around in confined spaces. We tracked the positions (supplementary materials and fig. S10) of the robots in the main tunnel area (i.e., excluding the excavation site) to generate space-time overlap maps of robot positions (see Fig. 4, B to D), which give visual insight into robot flow during excavation.

We first examined systematically how excavation performance changed as numbers of robots increased for our active protocol (as in Fig. 1C and fig. S6), which assigned equal work “desire” to all diggers: After soil deposition, each robot immediately returned to the tunnel to excavate. Despite constraints on maneuverability, sensing, and morphology, the robophysical experiments demonstrated qualitatively similar performance to the ants and the computational and theoretical models. For example, measurement of the average flow rate,  $\bar{q}$ , of successful excavators (which we quantify here as the number of deposits per minute) revealed that excavation performance increased with an increasing number of robots in the trial ( $N$ ) until the system became sufficiently crowded (Fig. 4E).

To characterize how clustering led to performance degradation in the active protocol, we measured the frequency of cluster occurrences, denoted  $I_c$ . Here we defined clusters as groups of robots of number  $N_c$ , whose center positions were within a robot length of each other (supplementary materials). Such clusters occurred most frequently at the excavation site (histograms in Fig. 4, B to D), yielding phase separation (29) in the system, whereby a portion of robots were jammed at high density, whereas others moved smoothly through the tunnel at low density.

As in (9), we also measured the characteristic “relaxation” times for clusters using a tool from the study of glassy systems, the density overlap correlation function  $Q(\tau)$ .  $Q(\tau)$  compares the spatial overlap of a cluster at a specific time to the overlap of the cluster’s original lateral segment at a later time,  $\tau$ . Assuming a one-dimensional tunnel, we calculated the spatial overlap of robots by tracking their centroid position laterally (along the tunnel) and assigning intensity potentials in space, summing overlapping potentials of adjacent robots (fig. S12A). From these curves (Fig. 4G, inset; fig. S12; and supplementary materials), we calculated the relaxation time,  $\tau^*$ , for clusters of different  $N_c$  by fitting a stretched exponential function,  $Q_f(\tau) = \exp\left\{-\left(\frac{\tau}{\tau^*}\right)^\beta\right\}$ , to a  $Q$  curve averaged over clusters of the same  $N_c$ , where  $\beta$  is a fitting parameter that is of order unity.

The relaxation time analysis highlighted how sufficient numbers of active robots ( $N = 4$ ) resulted in clustering cascades. For example,  $N_c = 2$  robot clusters could be sufficiently difficult to resolve before a third robot joined the cluster, which in turn led to catastrophic  $N_c = 4$  robot jams that spanned the tunnel width. Such clogs were then difficult to resolve with the

robots’ limited sensory and motor capabilities and were likely exacerbated by the robots’ rigid oblong shape (Fig. 4F). A sharp increase in  $\tau^*$  for clusters with a linear density,  $\bar{\lambda}_c = N_c/L_c$ , where  $L_c$  is cluster length in body lengths, greater than unity (multirobot clusters) during four-robot active trials (Fig. 4H, green curve) revealed how this cascading scenario is reminiscent of glassy arrest in particulate systems (30, 31).

To discover how the strategies of idleness distributions and reversals affected clustering and traffic dynamics in the robots (movie S5), we implemented two protocols inspired by the biological observations and theoretical models. As in the CA model, in the Lorenz protocol (fig. S8), we implemented an unequal probability to enter the tunnel derived from experimental ant workload inequalities. We also implemented a separate robot reversal protocol (fig. S7), which produced selective retreats, whereby the robots were programmed to immediately resume excavation after deposition but reversed after not successfully reaching the excavation site within a given time. These strategies led to different excavation performances as  $N$  increased; but most importantly, both strategies outperformed the active protocol at  $N = 4$  (Fig. 4E).

The relaxation times and cluster analysis revealed the mechanisms by which the different protocols mitigated clogging, particularly in the distinct ways in which they reduced the duration of clusters and thus optimized the average occupancy of excavators, thereby improving traffic flow. For trials with up to three robots, all strategies produced a relatively low  $\tau^*$  (Fig. 4G) and frequency of cluster occurrence regardless of the number of robots in a cluster. However, for  $N = 4$ , the Lorenz and reversal protocols mitigated the clogging effects associated with the aggressive excavation in the active protocol.

The Lorenz and reversal protocols provided distinct forms of mitigating the catastrophic cascades of clogs found in the active protocol: Unequal workload distributions reduce the occurrence of clusters, and selective retreating limits the duration of clogs. Selective retreating in the reversal strategy limited the duration of clogs. Thus, instead of the glass-forming characteristics of active robots, clusters dissolved after some time, yielding low  $\tau^*$  (Fig. 4, G and H). The unequal workload distributions of the Lorenz strategy reduced the occurrence of clusters, especially the highest-density four-robot clusters (Fig. 4I), where glasslike clog formation is most likely to occur, resulting in fewer catastrophic clogs at the excavation site. We found similar evidence for clog mitigation in the analysis of clusters in the CA model (fig. S26 and supplementary materials), whereby clog mitigation was further found to be most effective when both strategies (reversals and unequal entrance probabilities) were used in combination.

To close, we return to the traffic aspects of the confined system: As in theory, traffic flow of robotic ants (which dominates the excavation performance) was maximal at an intermediate occupancy of excavators,  $\bar{\lambda} = N_T/W_T$ , where  $N_T$  is

the number of robots in the main tunnel area averaged across all frames of video and  $W_T$  is the width of the tunnel, followed by a gradual decline at higher  $\bar{\lambda}$  (fig. S11A). However, unlike the theoretical models, peak flow rate in robotic systems occurred at a  $\bar{\lambda}$  of approximately 0.25, which corresponds to less than one robot traversing the tunnel at a time, despite a tunnel width of about three robot widths, or 1.5 robot lengths.

We hypothesize that the underperformance of our robots relative to the biological and theoretical systems is a consequence of our robots’ limited mobility in confined spaces, indicating that deformable bodies (32) and novel locomotor mechanisms (12) will be important in confined real-world robot collectives. That said, given these strategies are robust to the vagaries of real-world interactions, we posit that other engineered systems—including robot swarms in disaster rubble, nanorobots surging through the bloodstream (33), and task-capable active materials (3)—could benefit from simple strategies that involve labor inequality, particularly in creative combinations (34).

## REFERENCES AND NOTES

1. A. Okubo, *Adv. Biophys.* **22**, 1–94 (1986).
2. M. Brambilla, E. Ferrante, M. Birattari, M. Dorigo, *Swarm Intell.* **7**, 1–41 (2013).
3. M. Marchetti et al., *Rev. Mod. Phys.* **85**, 1143–1189 (2013).
4. D. Helbing, *Rev. Mod. Phys.* **73**, 1067–1141 (2001).
5. N. C. Dantton, L. Turner, S. Rojevsky, H. C. Berg, *Biophys. J.* **98**, 2082–2090 (2010).
6. E. Méhes, T. Vicsek, *Integr. Biol. (Camb.)* **6**, 831–854 (2014).
7. A. Garcimartín et al., *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **91**, 022808 (2015).
8. L. Giomi, N. Hawley-Weld, L. Mahadevan, *Proc. R. Soc. London A Math. Phys. Sci.* **469**, 20120637 (2013).
9. N. Gravish, G. Gold, A. Zangwill, M. A. D. Goodisman, D. I. Goldman, *Soft Matter* **11**, 6552–6561 (2015).
10. D. M. Gordon, *Nature* **380**, 121–124 (1996).
11. D. Cassill, W. R. Tschinkel, S. B. Vinson, *Insectes Soc.* **49**, 158–163 (2002).
12. N. Gravish, D. Monastanova, M. A. Goodisman, D. I. Goldman, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 9746–9751 (2013).
13. D. Monastanova et al., *J. Exp. Biol.* **218**, 1295–1305 (2015).
14. V. Linevich, D. Monastanova, D. I. Goldman, *Artif. Life Robot.* **21**, 460–465 (2016).
15. P. Tenczar, C. C. Lutz, V. D. Rao, N. Goldenfeld, G. E. Robinson, *Anim. Behav.* **95**, 41–48 (2014).
16. E. J. Robinson, O. Feinerman, N. R. Franks, *Proc. Biol. Sci.* **276**, 4373–4380 (2009).
17. D. M. Gordon, *Cell Syst.* **3**, 514–520 (2016).
18. V. Fourcassié, A. Dussutour, J. L. Deneubourg, *J. Exp. Biol.* **213**, 2357–2363 (2010).
19. A. John, A. Schadschneider, D. Chowdhury, K. Nishinari, *J. Theor. Biol.* **231**, 279–285 (2004).
20. N. Gravish et al., *J. R. Soc. Interface* **9**, 3312–3322 (2012).
21. C. R. Reid et al., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15113–15118 (2015).
22. C. Gershenson, D. Helbing, *Complexity* **21**, 9–15 (2015).
23. K. Nagel, M. Schreckenberg, *J. Phys. I* **2**, 2221–2229 (1992).
24. R. Kühne, in *Highway Capacity and Level of Service: Proceedings of the International Symposium on Highway Capacity*, Karlsruhe, 24–27 July 1991, U. Brannolte, Ed. (CRC Press, 1991).
25. H.-S. Kuan, M. D. Betterton, *Phys. Rev. E* **94**, 022419 (2016).
26. J. Aguilar et al., *Rep. Prog. Phys.* **79**, 110001 (2016).
27. M. J. Mataric, in *From Animals to Animals 2: Proceedings of the Second International Conference on Simulation of Adaptive Behavior*, J.-A. Meyer, H. L. Roitblat, S. W. Wilson, Eds. (MIT Press, 1993), pp. 432–441.
28. U. Borrmann, L. Wang, A. D. Ames, M. Egerstedt, *IFAC-PapersOnLine* **48**, 68–73 (2015).
29. G. S. Redner, M. F. Hagan, A. Baskaran, *Phys. Rev. Lett.* **110**, 055701 (2013).



30. D. I. Goldman, H. L. Swinney, *Phys. Rev. Lett.* **96**, 145702 (2006).
31. H.-S. Kuan, R. Blackwell, L. E. Hough, M. A. Glaser, M. D. Betterton, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **92**, 060501 (2015).
32. K. Jayaram, R. J. Full, *Proc. Natl. Acad. Sci. U.S.A.* **113**, E950–E957 (2016).
33. S. Li *et al.*, *Nat. Biotechnol.* **36**, 258–264 (2018).
34. W. Liu, A. F. Winfield, J. Sa, J. Chen, L. Dou, *Adapt. Behav.* **15**, 289–305 (2007).

# ACKNOWLEDGMENTS

The authors thank W. Gardner, Griffin Botanical Garden, and Chattahoochee-Oconee National Forest for giving us permission for ant collection. We would also like to acknowledge N. Gravish for insight and fruitful discussions; M. Kingsbury and L. Chen for

assistance in magnetic particle construction; R. Kutner, R. Srivastava, and J. Logan for their help with video analysis; and N. Conn for help with ant collection. H.-S.K. thanks the Max Planck Institute for the Physics of Complex Systems for providing computing resources. **Funding:** The authors acknowledge the support of National Science Foundation grants NSF PoLS-0957659, PHY-1205878, and DMR-1551095, as well as ARO grant W911NF-13-1-0347, the National Academies Keck Futures Initiative, and the Dunn Family Professorship (to D.I.G.). **Author contributions:** B.D. and D.M. collected the raw data for the ant experiments. W.S. and D.M. developed and performed the CA simulations. H.-S.K. and M.D.B. developed and analyzed the OAT model. V.L. constructed and performed the ant robot experiments, and J.A. tracked and analyzed the robot experiment data. All authors contributed to the preparation of the manuscript and were involved in the interpretation of results. **Competing interests:** The

authors declare that they have no competing interests. **Data and materials availability:** Data are available on SMARTech at <https://smartechn.gatech.edu/>.

# SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/361/6403/672/suppl/DC1](http://www.sciencemag.org/content/361/6403/672/suppl/DC1)  
Materials and Methods  
Figs. S1 to S26  
Tables S1 to S4  
References (35–41)  
Movies S1 to S5

6 April 2017; resubmitted 17 January 2018  
Accepted 14 June 2018  
10.1126/science.aan3891

## HIGH-PRESSURE PHYSICS

# Insulator-metal transition in dense fluid deuterium

Peter M. Celliers<sup>1\*</sup>, Marius Millot<sup>1</sup>, Stephanie Brygoo<sup>2</sup>, R. Stewart McWilliams<sup>3</sup>, Dayne E. Fratanduono<sup>1</sup>, J. Ryan Rygg<sup>1,4</sup>, Alexander F. Goncharov<sup>5</sup>, Paul Loubeyre<sup>2</sup>, Jon H. Eggert<sup>1</sup>, J. Luc Peterson<sup>1</sup>, Nathan B. Meezan<sup>1</sup>, Sebastien Le Pape<sup>1</sup>, Gilbert W. Collins<sup>1,4</sup>, Raymond Jeanloz<sup>6</sup>, Russell J. Hemley<sup>7</sup>

Dense fluid metallic hydrogen occupies the interiors of Jupiter, Saturn, and many extrasolar planets, where pressures reach millions of atmospheres. Planetary structure models must describe accurately the transition from the outer molecular envelopes to the interior metallic regions. We report optical measurements of dynamically compressed fluid deuterium to 600 gigapascals (GPa) that reveal an increasing refractive index, the onset of absorption of visible light near 150 GPa, and a transition to metal-like reflectivity (exceeding 30%) near 200 GPa, all at temperatures below 2000 kelvin. Our measurements and analysis address existing discrepancies between static and dynamic experiments for the insulator-metal transition in dense fluid hydrogen isotopes. They also provide new benchmarks for the theoretical calculations used to construct planetary models.

The transformation of hydrogen from a molecular insulator to an atomic metal at high densities has been a longstanding focus in physics and planetary science (1). The unique quantum metallic properties of the low-temperature solid (i.e., below 300 K) have drawn sustained interest (2–5), and characterizing the transformation in the hot, dense fluid is crucial for understanding the internal structure and dynamics of giant planets (6), including the origin of their large magnetic fields (7). Numerous studies of the insulator-metal (IM) transition in dense fluid hydrogen, beginning with theoretical work five decades ago, predicted a first-order transition in the fluid (8–10) with a critical point at very high temperatures (~13,000 to 15,000 K) and 60 to 90 GPa. However, the first experimental work on the IM transition in the fluid, carried out using dynamic compression techniques, provided evidence for a continuous transition with metallic states reached in the pressure range  $P = 50$  to 140 GPa and temperatures  $T = 3000$  to 8000 K (11–13). More recent predictions (14–17) placed the critical point at a much lower temperature (~2000 K). This motivated several experimental studies using static diamond anvil cell (DAC) techniques (18–22) and dynamic compression (23) to probe the fluid

properties below 2000 K and up to several hundred GPa.

Dynamic compression can explore a broad range of thermodynamic paths with time-varying manipulations of the applied pressure and controlled reverberation of pressure waves through the sample. This includes probing the dense fluid at temperatures below 2000 K, for example, with an initial jump in pressure delivered by a shock wave followed by shock reverberation or gradual ramp compression. The first demonstration of this strategy was carried out on deuterium with a magnetic compression technique at the Z facility (23). The results showed strong optical absorption beginning in the range 100 GPa <  $P$  < 130 GPa, followed by weak fluctuating reflectance in the range 130 GPa <  $P$  < 300 GPa, and culminated in abrupt jumps to high reflectance near 300 GPa. Knudson *et al.* (23) attributed the absorption to band gap closure and determined that the reflectance jumps were associated with the first-order IM transition. The reflectance jumps occurred at higher pressures upon compression than upon decompression, plausibly as a result of thermal conduction. Meanwhile, improvements in static compression methods have allowed the exploration of the behavior of the fluid over part of this pressure-temperature ( $P$ - $T$ ) range (up to 170 GPa and >1800 K) (18–22, 24–26). Changes in optical properties from 120 to 170 GPa depending on temperature were attributed to the IM transition (18–20, 22), whereas other experiments suggest the persistence of a finite (~1 eV) band gap at similar conditions (21).

The IM transition is the subject of a number of continuing theoretical studies (14–17, 27–29) that consistently predict a discontinuous transition below a critical point near ~2000 K, but over a broad range of pressures. Density functional theory (DFT)-based calculations show a spread in the transition pressure spanning 150 GPa, arising from the sensitivity of the boundary to

the choice of exchange-correlation functional used and whether zero-point energy is accounted for (1, 16). Quantum Monte Carlo (QMC) calculations should provide improved bounds on the transition pressures (16, 17), although they disagree with a recent benchmarking experiment (30). Transition pressures for hydrogen and deuterium are expected to be different because of isotope effects, but with a small relative magnitude. The transition in deuterium from QMC simulations is 30 GPa higher than in hydrogen at 600 K, decreasing to 10 GPa higher at 1200 K (16). Despite experimental support for a first-order IM transition (19, 20, 22, 23), the critical point has not been experimentally identified. Furthermore, the broad discrepancies in the measured transition pressure (20, 22, 23) and character (20–23) have made resolving the differences between the theoretical models challenging.

We completed a series of five dynamic compression experiments at the National Ignition Facility (NIF) to probe the IM transition up to 600 GPa at temperatures ranging from 900 K to 1600 K. The experiments were carried out using 168 laser beams to deliver up to 300 kJ of ultraviolet light that drove a near-isentropic reverberation compression of a cryogenic liquid deuterium sample. We adjusted the time dependence of the laser delivery (pulse shape) to control the compression sequence imposed on the sample as a function of time. Line-imaging Doppler velocimetry recorded both the compression history and the evolution of the optical properties of the D<sub>2</sub> sample during the nanosecond compression process, using a probe laser operating at 660 nm.

In our experimental setup, the fluid deuterium sample is sandwiched between a copper piston and a LiF window and is viewed through the window by the diagnostic (Fig. 1). The upper half of the view shows signal from the light reflected at the D<sub>2</sub>-LiF interface where a 100-nm-thick aluminum film was deposited (position >0  $\mu$ m in top panels of Fig. 1, B and C). The lower half shows the signal from light reflected initially at the piston surface (double-passed through the transparent sample layer). Because the two reflecting interfaces move differently, the two halves of the field of view display different apparent velocities (different fringe phases) until the time when the light intensity on the sample side reaches a minimum. At later times, the apparent velocity on the sample side matches that of the Al-LiF interface (common fringe phase), indicating that the optical reflection has shifted from the piston surface to the D<sub>2</sub>-LiF interface.

The copper piston (lower half) started moving near  $t = 10$  ns when an initial weak shock was transmitted into the sample layer. We controlled the first shock strength for each experiment in order to follow different isentropic compression paths. The first shock strengths varied from 1.8 GPa to 3.4 GPa (table S1). After the first shock crossed the sample layer, it reverberated against the LiF window (upper half, 16 ns) and continued to reverberate between the piston and the window while the sample layer was compressed. After several reverberations, both the piston and

<sup>1</sup>Lawrence Livermore National Laboratory (LLNL), Livermore, CA 94550, USA. <sup>2</sup>CEA, DAM, DIF, F-91297 Arpajon, France.

<sup>3</sup>School of Physics and Astronomy and Centre for Science at Extreme Conditions, University of Edinburgh, Edinburgh EH9 3FD, UK. <sup>4</sup>Department of Mechanical Engineering, Physics and Astronomy and Laboratory for Laser Energetics, University of Rochester, Rochester, NY 14623, USA.

<sup>5</sup>Geophysical Laboratory, Carnegie Institution of Washington, Washington, DC 20015, USA. <sup>6</sup>Department of Earth and Planetary Science and Department of Astronomy, University of California, Berkeley, CA 94720, USA. <sup>7</sup>Institute of Materials Science and Department of Civil and Environmental Engineering, The George Washington University, Washington, DC 20052, USA.

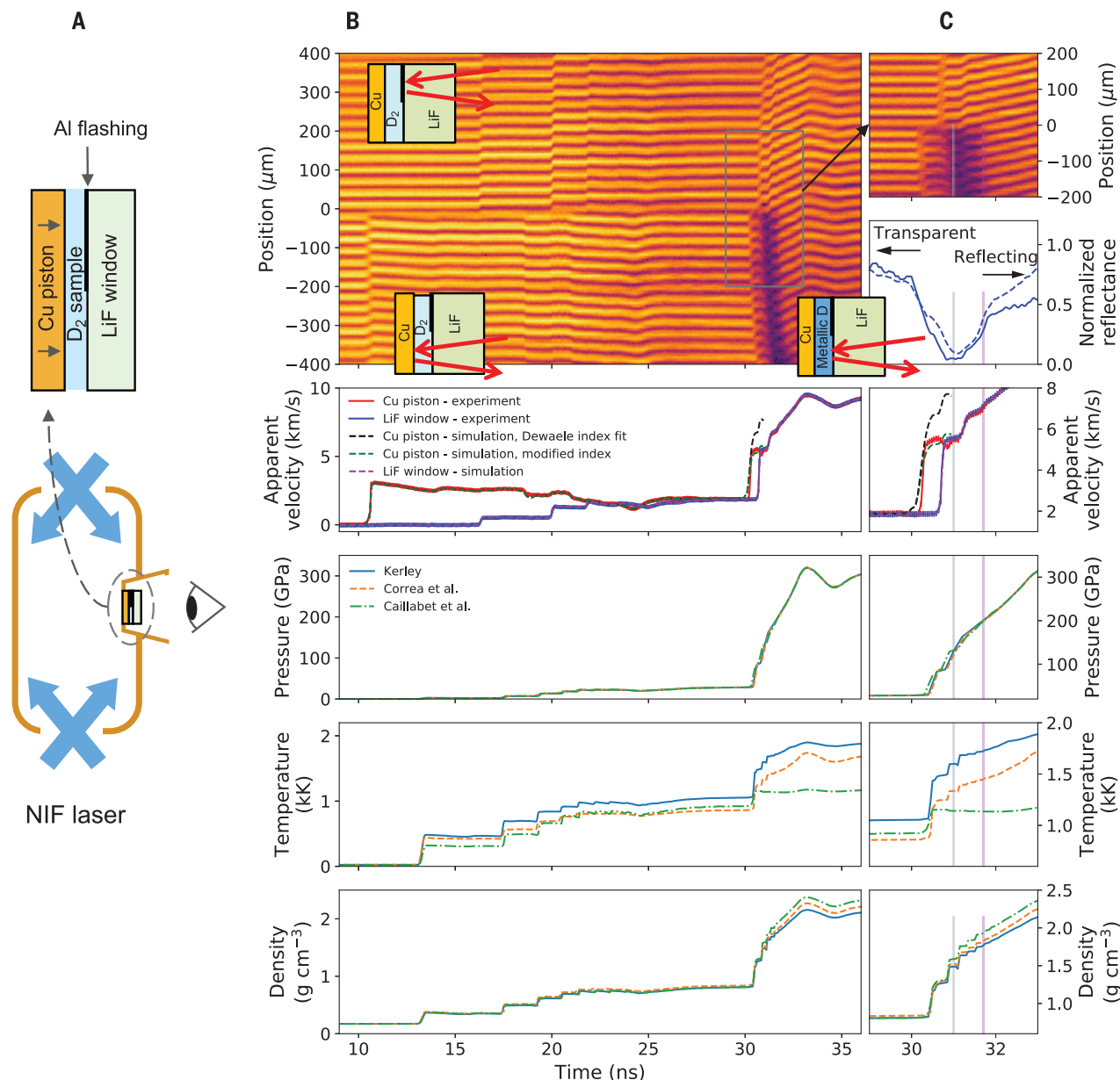
\*Corresponding author. Email: celliers1@llnl.gov

window interface approached a common apparent velocity by 27 ns. At this point, the sample was compressed to conditions estimated to be near density  $\rho \approx 0.8 \text{ g/cm}^3$ ,  $P \approx 20$  to 30 GPa, and  $T \approx 600$  to 900 K and was observed to be fully transparent. A second pressure wave arrived at the sample layer at 30 ns and initiated a sec-

ond set of reverberations, further pressurizing the sample to nearly 600 GPa in four experiments and to 215 GPa in one experiment. The duration of this second set of reverberations was shorter than the first, accomplishing much of the pressure increase in less than 3 ns. The transition from optically transparent to strongly

reflecting occurred during the second set of reverberations and is attributed to the IM transition in deuterium.

The thermodynamic path of the sample was inferred from hydrodynamic simulations constrained by the measured interface velocities (31), similar to previous work (23). Different



**Fig. 1. Schematic of target and experiment together with raw data.**

(A) A planar sample layer of liquid deuterium, 31  $\mu\text{m}$  thick, contained between a 70- $\mu\text{m}$ -thick copper piston and a 500- $\mu\text{m}$ -thick LiF window, is mounted onto the side of a hohlraum (brown in bottom sketch), driven with the NIF laser (blue arrows), and diagnosed with a line-imaging Doppler velocity interferometer (sensor positioned on right). (B) Example data record from experiment N150914-2. Fringe amplitude encodes reflectivity and fringe phase is proportional to the Doppler shift of the reflected light (piston and window interface velocities); the frame below shows apparent velocity signals extracted from the fringe phase and corresponding simulated velocities. At times  $30 < t < 31$  ns, simulations (black

dashed curve) with the extrapolated Dewaele *et al.* (36) refractive index fit do not match the observations and require a corrected index (green dashed curve) to match observations; additional frames below show pressure, temperature, and density at a point in the center of the sample estimated from simulations matched to the velocity data based on three different EOS models for deuterium. (C) Details of the information presented in (B) with magnified time axis. Vertical gray lines denote the transition from partially transparent to opaque; vertical magenta lines indicate the time when the reflectivity at the D/LiF interface exceeds 30%. The normalized reflectance shows independent measurements from the two detectors.



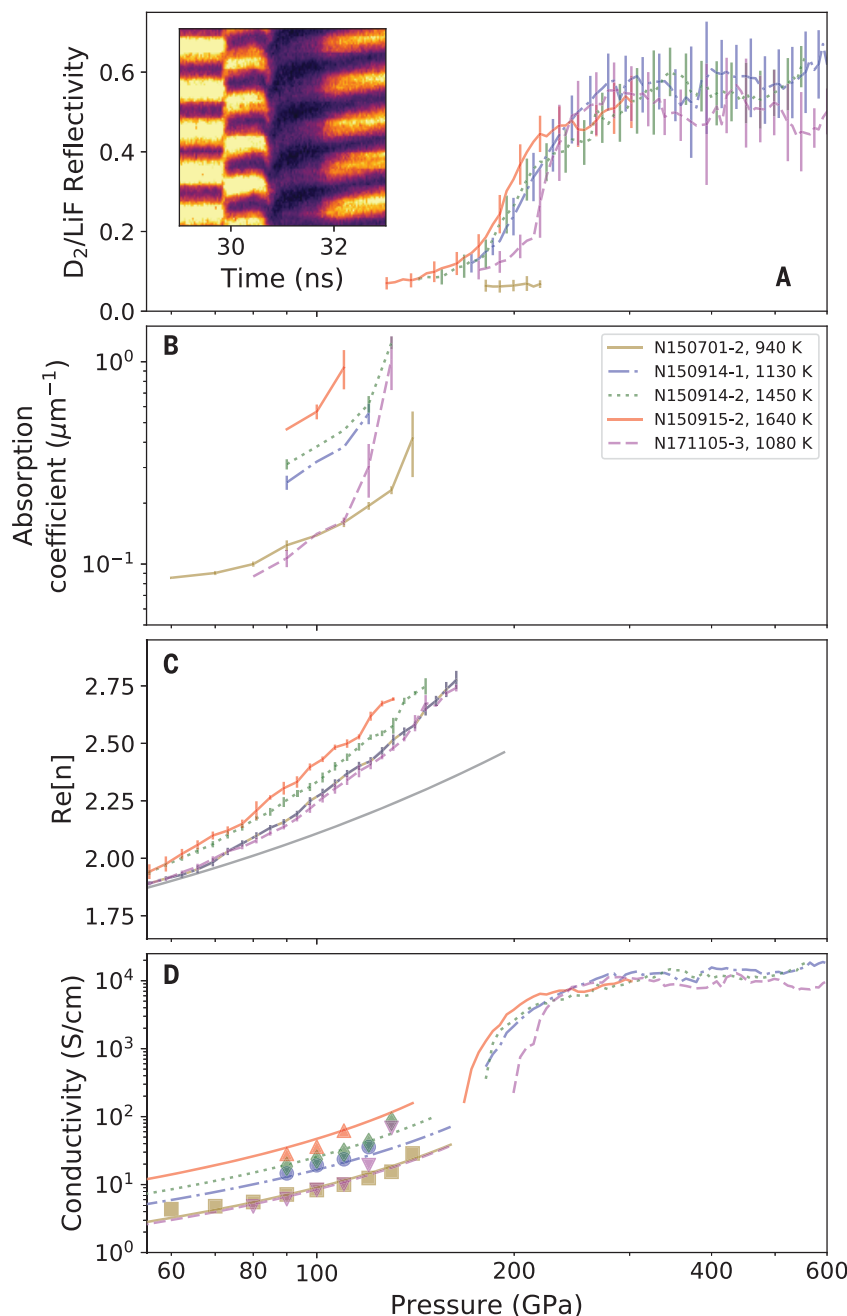
equation of state (EOS) models for deuterium (32–34) resulted in calculated pressures that are nearly the same to within 1% independent of the EOS model. We estimate that the pres-

sure along the compression path is accurate to  $\pm 3$  GPa (measurement error of the velocities), enabling determination of the metallization pressure to  $< 3\%$  accuracy (after accounting for

measurement error in the reflectance). The EOS models compare well with recent benchmarking experiments (13, 30, 35). However, none of the current EOS models represent the IM transition as a first-order phase transition (i.e., with a density jump and latent heat), and the range of temperature predictions among EOS models is large, leading to uncertainties of as much as  $\pm 260$  K near the IM transition. Density differs by less than  $\pm 6\%$  among the models.

Several optical signatures are associated with the transition from transparent to reflecting (Fig. 2). While the sample layer remained partially transparent, we determined the absorption coefficient up to a magnitude of about  $1 \mu\text{m}^{-1}$  (Fig. 2B), when the optical depth of the sample layer exceeded  $\sim 2$  and the layer became opaque. When the sample was opaque, light continued to be returned from the target because of the reflection from the  $\text{D}_2$ -LiF interface, with reflectance near 10%. While the sample was still partially transparent, we observed changes in the real part of the index of refraction that are consistent with band gap closure. We extracted the real part of the index of refraction (Fig. 2C) by fitting to the apparent velocity of the piston (observed through the deuterium sample) self-consistently with our hydrodynamic simulation (31) (Fig. 1). The real part of the index agrees closely with that of insulating hydrogen at lower pressure (36) but deviates from the low-pressure linear density dependence as the band gap decreases.

The gap energy, which we estimated from the refractive index (31), approaches  $\sim 2$  eV when the sample layer becomes opaque, consistent with the probe photon energy of 1.9 eV. Extrapolation (linear in density) suggests that for the fluid deuterium states achieved in our study, the band gap closes in the range  $200 \text{ GPa} < P < 250 \text{ GPa}$  (31) depending on the experiment (fig. S22B). The  $\sim 10\%$  reflectance we observed when the sample becomes opaque (Fig. 2A) is consistent with the expected reflectivity of the pressurized interface between materials of different indices of refraction,  $n_{\text{D}_2} \approx 2.8$  and  $n_{\text{LiF}} \approx 1.5$ , and represents a minimum reflectivity of insulating deuterium in this experiment. Further pressurization reveals a rapid increase in reflectivity at the  $\text{D}_2$ -LiF interface eventually reaching a saturation level near 55%. Reflectivity of  $\sim 30\%$  at the  $\text{D}_2$ -LiF interface corresponds to the minimum metallic electrical conductivity  $\sigma_{\text{DC}} \approx 2000 \text{ S/cm}$  (11, 37); values exceeding this indicate metallic behavior (31). From this measurement, we inferred that the IM transition occurs in the pressure range between approximately 150 GPa and 250 GPa. Although we did not observe a discontinuous change in reflectivity through this range for three of our experiments, finite temporal resolution in the measurements may have obscured a sharp transition; one experiment (Fig. 2A, magenta dashed curves and inset) recorded at higher time resolution revealed a sharp reflectivity increase. In all cases, the reflectivity increase steepened noticeably near 200 GPa. Of the four experiments that reached 600 GPa



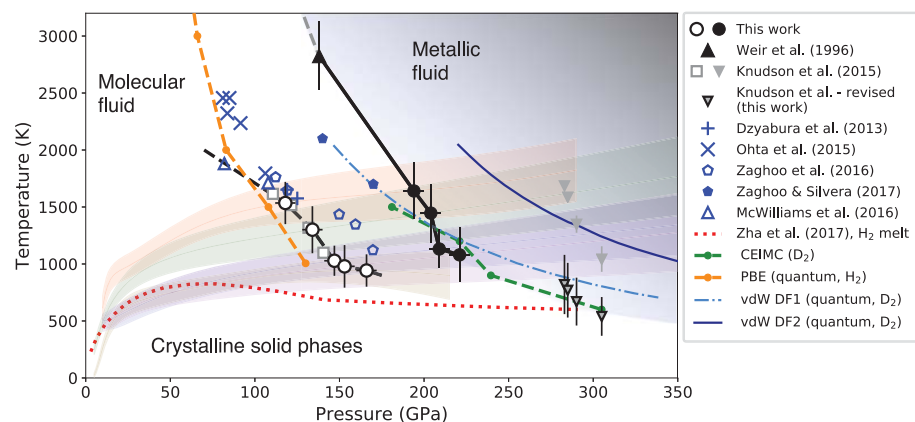
**Fig. 2. Optical signatures of the insulator-metal transition.** Curves in all frames are matched for the five experiments following isentropes that correspond to the colored  $P$ - $T$  bands in Fig. 3. (A) Reflectivity at the  $\text{D}_2$ -LiF interface as a function of pressure, after the  $\text{D}_2$  layer becomes opaque. Inset: Raw data from experiment N171105-3, showing a sharp reflectivity jump just before  $t = 32$  ns. (B) Absorption coefficient; legend indicates experiment number and  $T$  along isentrope near  $P = 200$  GPa. Colors and line types apply to all panels. (C) Index of refraction from simulations matched to the observed Doppler shifts and, for comparison, extrapolation of the Dewaele *et al.* (36) refractive index fit to higher density (solid gray curve). (D) For  $P < 150$  GPa, optical (AC) conductivity as extracted by combining the data in (B) and (C) (symbols with colors matched to curves) and corresponding Lorentz model fits (curves); for  $P > 150$  GPa, the DC conductivity inferred from the reflectivity data in (A) and the Smith-Drude model with  $\tau = 0.075$  fs (31). Error bars in (A) to (C) represent SD of multiple measurements collected in a set of velocity bins and mapped to the pressure scale.

peak pressure, we recorded some of the reflectivity data during pressure relaxation; unlike prior experiments (23), there is no evidence for different reflectivity signals during increasing ( $dP/dt > 0$ ) and decreasing ( $dP/dt < 0$ ) pressurization.

We can fit our results to simple models of the evolution of the optical properties through the transparent, optically absorbing, and optically reflecting regimes. The absorption has a steep pressure dependence and increases with temperature (Fig. 2B), most likely as a result of disorder in the material. In the partially transparent regime, we extracted the optical conductivity (Fig. 2D, symbols) directly from our data, using the expression  $\sigma(\omega) = n\omega c$  (Gaussian units) combining  $n$  (Fig. 2C) and the absorption coefficient  $\alpha$  (Fig. 2B); here,  $c$  is the speed of light. A Lorentz optical model incorporating a density-dependent oscillator frequency matches these data (Fig. 2D, curves below 150 GPa). We estimated the DC conductivity above the transition (Fig. 2D,  $P > 150$  GPa) with the Smith-Drude model (21, 38) evaluated with a fixed relaxation time  $\tau = 0.075$  fs and variable backscatter parameter matched to the optical reflectivity (Fig. 2A and fig. S25). The conductivities determined here exhibit trends similar to those calculated theoretically (14).

We plotted the optical absorption and reflectivity signatures along the calculated temperature versus pressure ( $T$ - $P$ ) paths extracted from the simulation models (Fig. 3), referred to by their  $T$  near 200 GPa (table S1). Our observations of the onset of optical absorption (Fig. 3, open circles) are in good agreement with the results of McWilliams *et al.* (21) and Knudson *et al.* (23) (Fig. 3, black dashed line). These observations, along with the extrapolated band gap energy as a function of  $P$  (fig. S22B), imply that the band gap remains finite leading up to the IM transition. These conclusions differ from those of several DAC measurements performed at similar conditions (18–20, 22). The DAC measurements using laser heating produced plateaus in temperature with increasing laser power that were interpreted as a signature of the metallization transition (18–20, 22). However, the heating plateaus in the static experiments are better correlated with the onset of absorption in our experiment rather than metallization, consistent with alternative interpretations of those data (21, 23, 39). The observation of absorption well below the point of band gap closure is also found in the solid at lower temperatures (40, 41).

We found saturation of the reflectivity above  $\sim 280$  GPa in the four experiments carried out to the highest pressures. In these datasets, the transition temperature exhibited a clear trend decreasing with pressure by  $\sim 20$  K/GPa. The lowest observation at  $T \sim 1100$  K showed the transition occurring at 221 GPa. One experiment at even lower  $T \sim 940$  K did not reveal reflectivity greater than 10% to the maximum  $P \sim 215$  GPa, but does not rule out the possibility of a transition at higher  $P$ . Pulsed-heating DAC experiments on  $H_2$  (20, 22) found reflectivity saturation at temperatures similar to our ex-



**Fig. 3. Phase diagram of  $H_2/D_2$  at high  $P$ - $T$  conditions.** Present results: Threshold where optical absorption coefficient exceeds  $\sim 1 \mu m^{-1}$  at 660 nm and the band gap is  $\sim 2$  eV (black open circles connected by thick dashed line); points where the D/LiF interface reflectivity exceeds  $R = 30\%$  (black solid circles), corresponding to the minimum metallic conductivity  $\sigma \approx 2 \times 10^3$  S/cm; vertical error bars are systematic (see below); horizontal error bars are estimated from a systematic component ( $\sim 1\%$  of  $P$ ) and random uncertainties in the velocity, reflectivity, and absorption data ( $\sim 6$  GPa). The thick black curve connects these points to the metallization transition,  $\sigma \approx 2 \times 10^3$  S/cm, identified by Weir *et al.* (solid black triangle, error bar estimated from systematic and random uncertainties) (11). The colored solid curves and associated bands show the compression paths of the present experiments; shading indicates estimated systematic uncertainty used to estimate the error bars in  $T$ . Previous experimental results on  $H_2$ : Ohta *et al.* ( $\times$ ) (19) and Dzyabura *et al.* ( $+$ ) (18), both from heating curve analyses of laser-heated DACs; Zaghou *et al.* (open pentagons) (20) from absorption and reflectance in laser-heated DAC; McWilliams *et al.* (open triangles) (21) determination of the onset of absorption from transient DAC optical transmission measurements; Zaghou and Silvera (solid pentagons) (22) reflectance saturation data; melting measurements of Zha *et al.* (dotted red curve) (24). Previous results on  $D_2$ : Knudson *et al.* (open squares) (23) absorption onset for 532-nm wavelength; Knudson *et al.* (inverted gray triangles) (23) reported IM transition and revised with our interpretation (inverted black-edged triangles); vertical error bar estimates are explained in (31). Selected theoretical curves reported by Pierleoni *et al.* (16) and Knudson *et al.* (23) are shown, including the CEIMC calculation for deuterium (dashed green curve) and DFT calculations with quantum corrections for the ions [vdW-DF1 for deuterium (dot-dashed blue curve), vdW-DF2 for deuterium (dark blue curve), and PBE for hydrogen (dashed orange curve)].

periments and at pressures that were lower by  $\sim 20$  GPa, which may be consistent with the expected isotope effect.

The steep increase in reflectivity with pressure that we observe near 200 GPa contrasts with the interpretation of the ramp compression experiment on deuterium carried out at the Z facility (23). Although the onset of optical absorption is consistent with our work, the subsequent weak reflectivity followed by jumps near 300 GPa (attributed to the IM transition) occurred at higher pressure than observed here and in other later studies (20, 22), whereas the low ( $< 2\%$ ) reflectivity of the  $D_2$ -LiF interface recorded at intermediate conditions is not consistent with either insulating or semiconducting deuterium and LiF under pressure (31, 36, 42, 43). Our data indicate onset of the IM transition at  $P \sim 200$  GPa; the optical reflectivity saturates to a constant value of  $\sim 55\%$  near 280 GPa and remains stable over hundreds of GPa, similar to the pressures where high reflectivity is displayed in the Z experiments. This shows that both experiments are probing similar states at onset and completion of the IM transition, but not at intermediate conditions.

We propose that the different results are related to the time scale of the Z experiments being two orders of magnitude longer, allowing turbulent lateral flows within the sample layer to be established during the observations. These effects lead to a different interpretation of the reflectivity jumps observed in the Z data, correlating these with completion rather than onset of the IM transformation (31). This requires a correction of the inferred transition temperatures from the Z-machine to lower values (table S3), owing to the latent heat of the IM transition. Although there are uncertainties in the temperature estimates in both experiments, we expect that those uncertainties are accentuated by the longer time scales, which give rise to multiple mechanisms of heat transport in the Z experiments. The much shorter time scales in our experiments preclude these from occurring, so that the temperature estimates for our experiments depend primarily on the EOS models in the molecular fluid phase.

We interpreted the onset of reflectivity above 10% as the beginning of pressure-induced molecular dissociation associated with the IM transition. The extent and  $P$ - $T$  range of the IM

transition are in approximate agreement with the DFT-based molecular dynamics calculations using the van der Waals (vdW-DF1) exchange-correlation functional (44) reported in Knudson *et al.* (23). Our observations and analyses are also consistent with the coupled electron-ion Monte Carlo method (CEIMC) calculations of Pierleoni *et al.* (16) and with the quantum Monte Carlo calculations of Mazzola *et al.* (17). They are about 100 GPa higher than DFT calculations based on the Perdew-Burke-Ernzerhof (45) (PBE) exchange-correlation functional (15, 29) and about 50 GPa lower than DFT calculations based on the van der Waals vdW-DF2 functional (23, 46). In the saturation regime, the estimated  $\sigma_{DC} \approx 10^4$  S/cm is in good agreement with the recent estimates from lower-pressure experiments (22) and with calculations (14).

Our interpretation of the Z experiments does not alter the conclusion that the abrupt reflectivity jumps observed in those experiments are evidence of a first-order IM transition. The revised  $T \approx 700$  K is well below the predicted critical-point temperature  $T_{CP}$ , whereas one of our experiments provides evidence for first-order behavior up to 1100 K. The extrapolated band gap energies (linear in density; fig. S22B) reach closure at higher  $P$  than the corresponding observed transition  $P$ , indicating that the gap closes nonlinearly and could indicate first-order behavior at temperatures as high as 1640 K. Therefore, although  $T_{CP}$  is not yet precisely identified, it is bounded in the range  $3000 \text{ K} > T_{CP} > 1100 \text{ K}$ , as indicated by our data and by continuous transformations observed in prior experiments to higher temperature (11).

Our revision of the IM transition data from the Z experiments produces points that lie close to and even below previous melting-line studies on hydrogen (24, 26). A straight line connecting our data to the revised Z data has a slope of  $-5 \text{ K/GPa}$ . This invites speculation as to whether a thermodynamic triple point with coexisting crystalline solid, insulating fluid, and liquid metal phases is located near 600 K and 300 GPa. If so, at higher pressures the insulating crystalline solid would transform directly to the liquid metal. Because the melting line of deuterium has not been determined and because quantum effects are important, the existence and location of such a triple point remains an open question.

Demixing in dense H-He fluid mixtures, an important process expected to occur inside giant planets (47, 48), is closely connected with the IM transition (49). Recent state-of-the-art calculations of H-He demixing are all based on ab initio calculations with increasing levels of sophistication (49–55). The two most recent calculations of H and He EOS for planetary models are based on DFT molecular dynamics (56, 57) with the PBE exchange-correlation functional, and until this year, ab initio-based demixing studies (53, 54)

were also based on PBE. The emerging theoretical (16, 17) and experimental (23) evidence that PBE underestimates the IM transition pressure motivated Schöttler and Redmer (55) to examine H-He demixing with the vdW-DF1 functional. They showed a clear shift of the demixing boundary to higher pressure and lower temperature relative to PBE-based calculations. Thus, although the first-order IM transition occurs at much lower temperature than the Jupiter and Saturn isentropes, it plays an important role as a benchmarking feature to validate the functional on which the H-He demixing calculations are based. Our data suggest that the vdW-DF1 functional is the current best choice.

## REFERENCES AND NOTES

- J. M. McMahon, M. A. Morales, C. Pierleoni, D. M. Ceperley, *Rev. Mod. Phys.* **84**, 1607–1653 (2012).
- N. W. Ashcroft, *Phys. Rev. Lett.* **21**, 1748–1749 (1968).
- E. G. Browman, Y. Kagan, A. Kholas, *Sov. Phys. JETP* **35**, 783 (1972).
- E. Babaev, A. Sudbø, N. W. Ashcroft, *Nature* **431**, 666–668 (2004).
- R. P. Dias, I. F. Silvera, *Science* **355**, 715–718 (2017).
- N. Nettelmann, A. Becker, B. Holst, R. Redmer, *Astrophys. J.* **750**, 52 (2012).
- J. E. P. Connerney *et al.*, *Space Sci. Rev.* **213**, 39–138 (2017).
- G. É. Norman, A. N. Starostin, *J. Appl. Spectrosc.* **13**, 965–967 (1970).
- W. Ebeling, W. Richert, *Phys. Lett. A* **108**, 80–82 (1985).
- D. Saumon, G. Chabrier, *Phys. Rev. Lett.* **62**, 2397–2400 (1989).
- S. T. Weir, A. C. Mitchell, W. J. Nellis, *Phys. Rev. Lett.* **76**, 1860–1863 (1996).
- P. M. Celliers *et al.*, *Phys. Rev. Lett.* **84**, 5564–5567 (2000).
- P. Loubeyre *et al.*, *Phys. Rev. B* **86**, 144115 (2012).
- W. Lorenzen, B. Holst, R. Redmer, *Phys. Rev. B* **82**, 195107 (2010).
- M. A. Morales, C. Pierleoni, E. Schwegler, D. M. Ceperley, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12799–12803 (2010).
- C. Pierleoni, M. A. Morales, G. Rillo, M. Holzmann, D. M. Ceperley, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 4953–4957 (2016).
- G. Mazzola, R. Helled, S. Sorella, *Phys. Rev. Lett.* **120**, 025701 (2018).
- V. Dzyabura, M. Zaghou, I. F. Silvera, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 8040–8044 (2013).
- K. Ohta *et al.*, *Sci. Rep.* **5**, 16560 (2015).
- M. Zaghou, A. Salamat, I. F. Silvera, *Phys. Rev. B* **93**, 155128 (2016).
- R. S. McWilliams, D. A. Dalton, M. F. Mahmood, A. F. Goncharov, *Phys. Rev. Lett.* **116**, 255501 (2016).
- M. Zaghou, I. F. Silvera, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11873–11877 (2017).
- M. D. Knudson *et al.*, *Science* **348**, 1455–1460 (2015).
- C. S. Zha, H. Liu, J. S. Tse, R. J. Hemley, *Phys. Rev. Lett.* **119**, 075302 (2017).
- N. Subramanian, A. F. Goncharov, V. V. Struzhkin, M. Somayazulu, R. J. Hemley, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 6014–6019 (2011).
- R. T. Howie, P. Dalladay-Simpson, E. Gregoryanz, *Nat. Mater.* **14**, 495–499 (2015).
- W. J. Nellis, A. A. Louis, N. W. Ashcroft, *Philos. Trans. R. Soc. A* **356**, 119–138 (1998).
- S. Scandolo, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3051–3053 (2003).
- I. Tamblyn, S. A. Bonev, *Phys. Rev. Lett.* **104**, 065702 (2010).
- M. D. Knudson, M. P. Desjarlais, *Phys. Rev. Lett.* **118**, 035501 (2017).
- See supplementary materials.
- G. I. Kerley, “Equations of State for Hydrogen and Deuterium” (SAND2003-3613, Sandia National Laboratories, 2003).
- L. Caillabet, S. Mazevet, P. Loubeyre, *Phys. Rev. B* **83**, 094101 (2011).
- A. A. Correa, L. X. Benedict, M. A. Morales, P. A. Sterne, J. I. Castor, E. Schwegler, arXiv:1806.01346 (2017).
- S. Bryggo *et al.*, *J. Appl. Phys.* **118**, 195901 (2015).
- A. Dewaele, J. H. Eggert, P. Loubeyre, R. Le Toullec, *Phys. Rev. B* **67**, 094112 (2003).
- W. J. Nellis, S. T. Weir, A. C. Mitchell, *Phys. Rev. B* **59**, 3434–3449 (1999).
- N. V. Smith, *Phys. Rev. B* **64**, 155106 (2001).
- Z. M. Geballe, R. Jeanloz, *J. Appl. Phys.* **111**, 123518 (2012).
- H. Mao, R. J. Hemley, *Rev. Mod. Phys.* **66**, 671–692 (1994).
- P. Loubeyre, F. Occelli, R. LeToullec, *Nature* **416**, 613–617 (2002).
- D. E. Fratanduono *et al.*, *J. Appl. Phys.* **109**, 123521 (2011).
- P. A. Rigg, M. D. Knudson, R. J. Scharrif, R. S. Hixson, *J. Appl. Phys.* **116**, 033515 (2014).
- M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, B. I. Lundqvist, *Phys. Rev. Lett.* **92**, 246401 (2004).
- J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
- K. Lee, É. D. Murray, L. Kong, B. I. Lundqvist, D. C. Langreth, *Phys. Rev. B* **82**, 081101 (2010).
- E. E. Salpeter, *Astrophys. J.* **181**, L83 (1973).
- D. J. Stevenson, E. E. Salpeter, *Astrophys. J. Suppl. Ser.* **35**, 239–261 (1977).
- W. Lorenzen, B. Holst, R. Redmer, *Phys. Rev. Lett.* **102**, 115701 (2009).
- J. E. Klepeis, K. J. Schafer, T. W. Barbee 3rd, M. Ross, *Science* **254**, 986–989 (1991).
- O. Pfaffenzeller, D. Hohl, P. Ballone, *Phys. Rev. Lett.* **74**, 2599–2602 (1995).
- J. Vorberger, I. Tamblyn, B. Militzer, S. A. Bonev, *Phys. Rev. B* **75**, 024206 (2007).
- M. A. Morales *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 1324–1329 (2009).
- M. A. Morales, S. Hamel, K. Caspersen, E. Schwegler, *Phys. Rev. B* **87**, 174105 (2013).
- M. Schöttler, R. Redmer, *Phys. Rev. Lett.* **120**, 115703 (2018).
- B. Militzer, W. B. Hubbard, *Astrophys. J.* **774**, 148 (2013).
- A. Becker *et al.*, *Astrophys. J. Suppl. Ser.* **215**, 21 (2014).

## ACKNOWLEDGMENTS

We thank J. Kroll, the target fabrication team at LLNL, and the NIF operations crew for experimental support. We thank the anonymous referees for extensive and detailed comments and questions that resulted in an improved manuscript. **Funding:** This work was performed under the auspices of the U.S. Department of Energy by LLNL under contract DE-AC52-07NA27344, Lawrence Livermore National Security, LLC. This material is based on work supported by the Department of Energy National Nuclear Security Administration (DOE/NNSA) under grant DE-NA0001944, the University of Rochester, and the New York State Energy Research and Development Authority. Also supported by EPSRC under grant Ep/P024513/1 (R.S.M.); the Army Research Office (56122-CH-H), the National Natural Science Foundation of China (21473211), and the Chinese Academy of Science (YZ201524) (A.F.G.); DOE/NNSA (including DE-NA0003607) and the University of California (R.J.); and DOE/NNSA (DE-NA0002006, CDAC) (R.J.H.). **Author contributions:** R.S.M., R.J., R.J.H., G.W.C., P.M.C., M.M., S.B., P.L., J.R.R., A.F.G., and J.H.E. conceived the experiments; J.L.P., N.B.M., P.M.C., M.M., D.E.F., and S.B. designed the experiments; P.M.C., M.M., S.L.P., S.B., D.E.F., and A.F.G. executed the experiments; P.M.C., S.B., M.M., and R.S.M. analyzed the data and performed post-shot simulations; P.M.C., R.J.H., R.S.M., M.M., and R.J. wrote the manuscript; and all authors reviewed and discussed the manuscript during preparation. **Competing interests:** All authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

## SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6403/677/suppl/DC1  
Materials and Methods  
Supplementary Text  
Figs. S1 to S34  
Tables S1 to S3  
References (58–108)

24 January 2018; accepted 31 May 2018  
10.1126/science.aar0970



## MEMBRANES

# 3D printed polyamide membranes for desalination

Maqsud R. Chowdhury<sup>1</sup>, James Steffes<sup>2</sup>, Bryan D. Huey<sup>2</sup>, Jeffrey R. McCutcheon<sup>1\*</sup>

Polyamide thickness and roughness have been identified as critical properties that affect thin-film composite membrane performance for reverse osmosis. Conventional formation methodologies lack the ability to control these properties independently with high resolution or precision. An additive approach is presented that uses electrospraying to deposit monomers directly onto a substrate, where they react to form polyamide. The small droplet size coupled with low monomer concentrations result in polyamide films that are smoother and thinner than conventional polyamides, while the additive nature of the approach allows for control of thickness and roughness. Polyamide films are formed with a thickness that is controllable down to 4-nanometer increments and a roughness as low as 2 nanometers while still exhibiting good permselectivity relative to a commercial benchmarking membrane.

The thin-film composite (TFC) membrane has served as the desalination industry's standard membrane for more than 30 years. During that time, this membrane has changed little. The composite structure comprises a polyester backing layer for mechanical support, a porous supporting polysulfone mid-layer cast through phase inversion, and an ultrathin, highly cross-linked polyamide film that is dense enough to separate salt ions from water but thin enough to have a low resistance to water transport. This polyamide layer is formed *in situ* onto the porous midlayer via interfacial polymerization. This approach relies on a reaction between an amine [*m*-phenylene diamine (MPD)] in an aqueous phase and an acid chloride [trimesoyl chloride (TMC)] in an organic phase. The immiscibility of the two phases permit the reaction to occur only at the phase boundary. Film growth is limited to the boundary and subsequently self-limits the reaction as reactants are blocked by the growing film. The result is a self-terminated, but uncontrolled, film growth with a thickness between 100 and 200 nm and a rough ridge-and-valley-like surface morphology (1–3). Although these membranes exhibit excellent permselectivity compared with any other desalination membrane, certain features of the film properties and its fabrication procedure are inherently limiting. The intrinsic roughness of these films have long been attributed to a high fouling propensity for reverse osmosis and nanofiltration processes (4, 5). Additionally, the thickness of the membrane, which is inversely proportional to its permeance, is relatively uncontrolled because the process simply self-terminates as the film forms. Last, the properties of the support layer surface—

including pore size, pore spacing, surface porosity, roughness, and surface chemistry—affect the interface between the two phases and thus the membrane performance in unpredictable ways (6–8).

A better polyamide desalination membrane should have the same permselective properties as those of existing membranes but also be tunable in each of these other properties. The thickness should be reduced to maximize permeance while still ensuring that the films are sufficiently robust so as to withstand necessary hydraulic pressures. The roughness should be minimized to lessen the likelihood that the membrane will foul and also improve cleaning efficiency. Last, the film properties should be decoupled from the substrate properties, allowing these selective films to be deployed on any type of substrate.

To better control thickness and roughness, Gu *et al.* used a molecular layer-by-layer approach to build polyamide layers onto ultrafiltration (UF) membranes. Using a polyelectrolyte layer to prime the surface of a porous substrate, a polyamide layer could be formed by molecular layer through a sequential interfacial polymerization method (9). Karan *et al.* used a sacrificial nanostrand layer as a support to form free-standing polyamide films with varying thickness and roughness for organic solvent nanofiltration applications with no demonstration of desalination performance (10). These methods and others (11–16) are complex and are unlikely to scale easily for commercial production.

Electrospray can be used to deposit monomers as nanoscale droplets that form polyamide onto a substrate. During electrospraying, liquid leaves a needle in the presence of a strong electric field. Coulombic repulsion forces the ejected droplets to disperse with diameters well below 1  $\mu\text{m}$  (Fig. 1, A and B) (17). This characteristic drew Fenn *et al.* to use the technique for mass spectrometry of large polar biomolecules (18, 19). Others followed by using the technique to make thin films (20–23), nanoparticles (24), or patterns (25–27). For our approach, we deposit individual monomers onto

a substrate, where they can subsequently polymerize on the surface.

The approach is illustrated in Fig. 1, A and B. The drum is grounded and connected to the two needles by means of a high-voltage dc power source that can generate up to 30 kV. The distance between the needle tip and drum is kept at 2 to 3 cm. Each needle extrudes one of the monomers in solution. MPD (in water) and TMC (in hexane) were kept at a molar ratio of 4:1 over a wide range of concentrations (table S1). A lipophilic ionic liquid was added to the organic phase in order to increase the electrical conductivity (fig. S2A). A variety of UF membrane substrates with different pore sizes (fig. S2B), pure water permeance (fig. S2C), and hydrophilicity were studied (fig. S2D and table S2). In each case, the substrate was first attached to the rotating drum (Fig. 1A). As monomer solutions emerged from the needle tips, they sprayed and deposited onto the collector surface and reacted upon contact with each other. To ensure coverage over the entire substrate, the needle stage traverses along the collector surface (Fig. 1B). A single pass over the collector surface is referred to as a “single scan.”

Films were printed on aluminum (Al) foil in order to demonstrate the ability to characterize polyamide films to find properties such as cross-link density, thickness, and mechanical properties. After printing, the films are transferred from the foil (fig. S3A) to any substrate or kept as a free-standing film (Fig. 1C). Having thicker films that can be manipulated by hand allows for easier characterization of film properties. This type of manipulation is difficult with conventional polyamide films because of their thinness, fragility, and integration into the supporting structure of typical TFC membranes. For example, determination of cross-link density of the polyamide film is typically done with x-ray photoelectron spectroscopy (XPS) (28). However, this method can be inaccurate because of surface roughness, insufficient sample size, and compositional heterogeneity with depth. Instead, manipulating a 1- $\mu\text{m}$ -thick polyamide into a thicker, crumpled form (fig. S4) allows us to use energy-dispersive x-ray spectroscopy, which penetrates far deeper into the sample and thus provides a better measurement of bulk polyamide composition. The cross-link density is found to be 83%, which is reasonable for a film made from MPD and TMC monomers (fig. S4) (10).

Films are also printed at various MPD and TMC concentrations (table S1) onto Al foil and then transferred to a silicon wafer for thickness measurement by using atomic force microscopy (AFM) (fig. S3B). Cross sections at the film edges (fig. S5) reveal the film profile with respect to the underlying planar substrate. Lower monomer concentrations not only result in a thinner polyamide film but also greater control of film thickness per scan. Polyamide films as thin as 20 nm were made based on five scans, indicating a mean thickness of just 4 nm per scan (Fig. 1D). Control of thickness per scan was notably consistent; linearity in film growth with an increasing number of scans is depicted in Fig. 1E.

<sup>1</sup>Department of Chemical and Biomolecular Engineering, University of Connecticut, Center for Environmental Sciences and Engineering, 191 Auditorium Road, Unit 3222, Storrs, CT 06269-3222, USA. <sup>2</sup>Department of Materials Science and Engineering, University of Connecticut, 97 North Eagleville Road, Unit 3136, Storrs, CT 06269-3136, USA.

\*Corresponding author. Email: jeffrey.mccutcheon@uconn.edu

Films of the same composition were also printed onto porous polymeric substrates in order to evaluate their thickness, surface morphology, roughness, desalination performance, and substrate independence. Cross-sectional transmission electron microscopy (TEM) images are shown in Fig. 1, F to I, and fig. S6. The polyamide layers printed on the three UF membrane substrates exhibit similar thicknesses (Fig. 1, F to H) as those printed on Al foils (Fig. 1D). We note repeatability in thickness from Fig. 1I, where five layers of polyamide film measuring  $15 \pm 3$  nm each are visible. This thickness per scan corresponds well to thickness per scan data captured on Al foil in Fig. 1D by means of AFM. We also confirm linearity in thickness with TEM images shown in fig. S6.

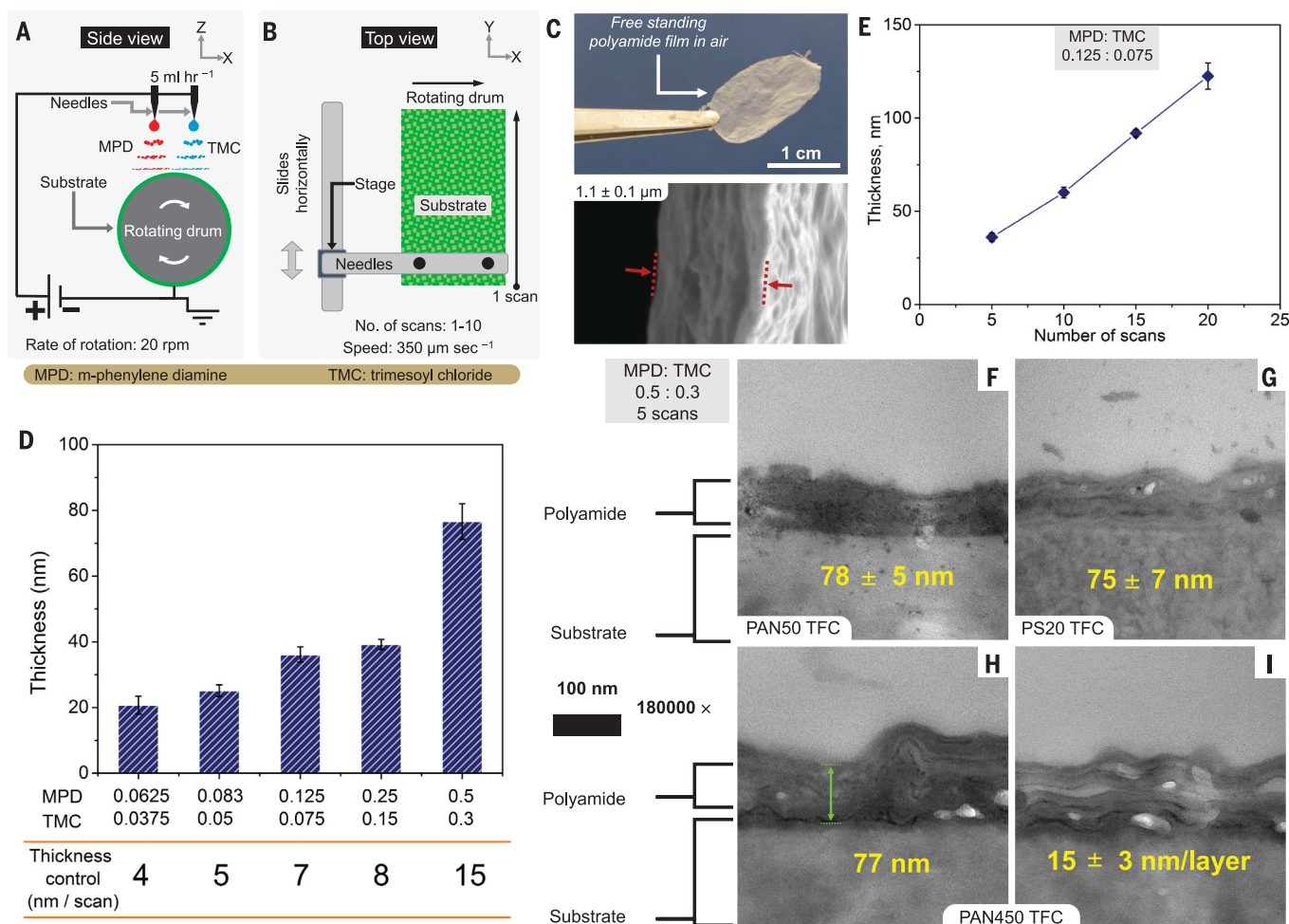
We examined the surface morphology of the polyamide films formed on polymeric substrates using scanning electron microscopy (SEM) (Fig. 2A

and figs. S7 and S8). Compared with the typical ridge-and-valley-like morphology of conventional polyamide films, such as the industry-standard Dow SW30XLE RO membrane (Fig. 2A), significantly smoother polyamide films are formed on all substrates at all monomer concentrations. These results are quantified by means of AFM analysis as shown in Fig. 2B. The root mean square (RMS) roughness increases with increasing monomer concentration (Fig. 2C) and the number of scans (Fig. 2D). For each monomer concentration, film roughness is similar among all of the substrates evaluated (fig. S9 and tables S3 and S4). The maximum roughness ( $40 \pm 4$  nm) is observed for the highest MPD:TMC concentration, 0.5:0.3 (Fig. 2C), when formed on the PAN450 UF substrate. However, even these roughest films exhibit less than half of the roughness of the Dow SW30XLE membrane (Fig. 2C, dotted orange overlay). The lowest concentrations of monomers

tested here yield films with roughness values of less than 2 nm and are indistinguishable from the substrate's roughness.

The desalination performance for all membranes tested are presented in Fig. 3A, where higher salt rejection and water permeance are desired. Using the SW30XLE as a control and for benchmarking purposes, six of our membranes had both higher rejection and water permeance (within the Fig. 3A gray rectangular overlay), and 30 are higher in one metric or the other. Although it was not the intent of this work to outperform an industry-standard membrane in conventional metrics of water permeance and salt rejection, these membranes can have tailorable thickness and substantially lower roughness while exhibiting comparable (or better, in some cases) performance.

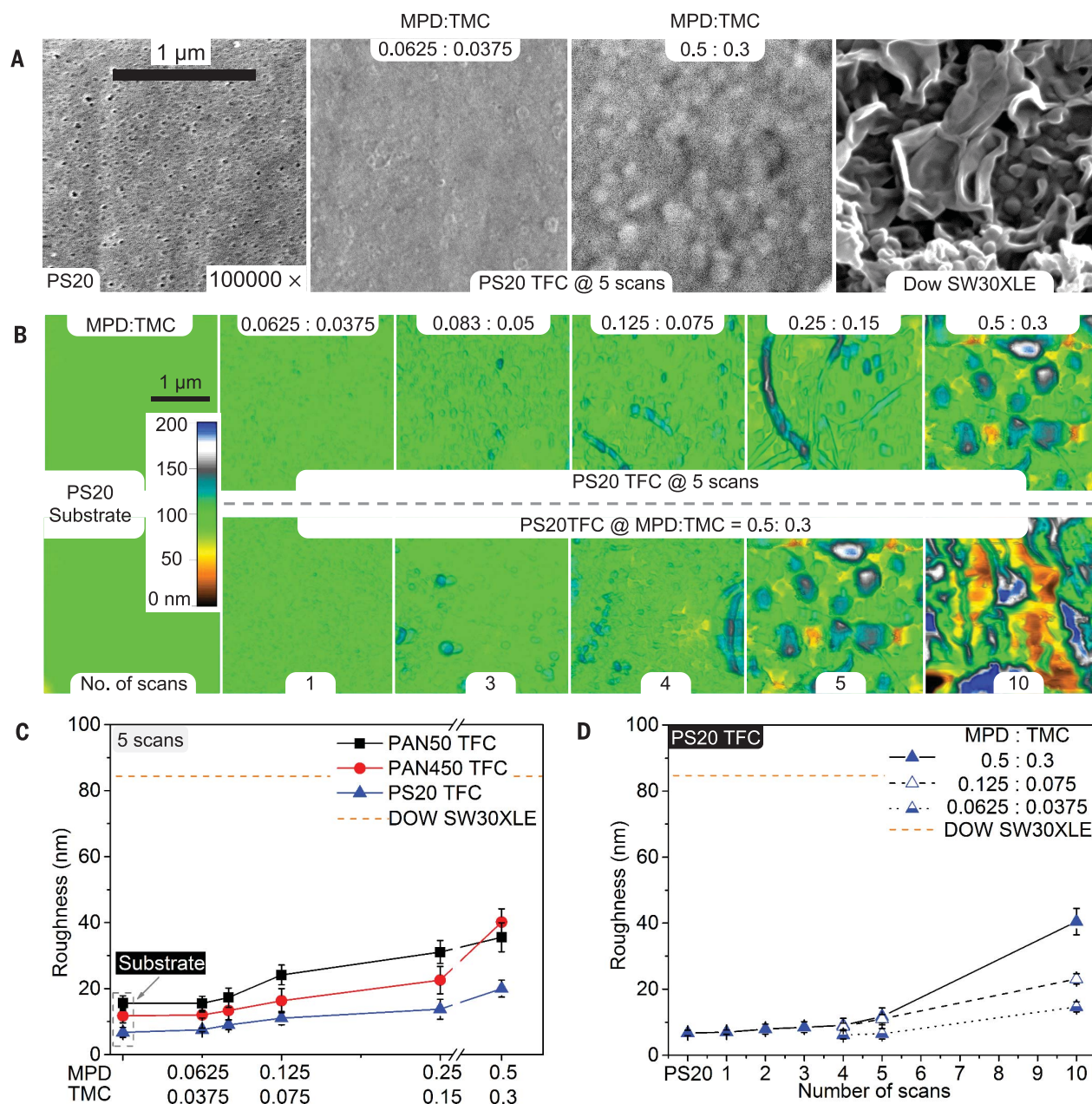
Water permeance (Fig. 3B) and salt rejection (Fig. 3C) are shown to have a strong dependency



**Fig. 1. Details of the electrospray process for printing substrate-independent polyamide films with thickness control.** (A) A side view of a schematic of the electrospray process. (B) The top view schematic shows the needles and a stage assembly that can move “horizontally” for uniform coatings on a rotated drum. A single sweep across the substrate is denoted as a single scan. (C) A free-standing polyamide film measuring  $1.1 \mu\text{m}$  thick in air, along with the cross-section from SEM. (D) Polyamide thickness as a function of MPD and TMC loading, including the corresponding thickness

per scan. (E) Polyamide thickness as a function of the number of scans at a MPD:TMC concentration ratio of 0.125:0.075. For characterization data presented in (C) to (E), the polyamide was prepared on an Al foil substrate and then separated according to fig. S3A. (F to I) Cross-section TEM of (F) PAN50, (G) PS20, and [(H) and (I)] PAN450 TFC membranes made with five scans and a MPD:TMC concentration ratio of 0.5:0.3. The displayed thickness and error represents 20 measurements from the images, except for (H), where only the thinnest region is measured.





**Fig. 2. Dependency of surface morphology and roughness on printing conditions.** (A) SEM image of TFC membranes at 100,000 $\times$  magnification for different concentrations of MPD and TMC. The underlying substrate and a Dow SW30XLE membrane are shown as controls. (B) A series of 3- by 3- $\mu\text{m}$  AFM topography images reveal increased surface roughness with the MPD:TMC concentration ratio, either consistently with five scans (top) or due to successive scans for the specific MPD:TMC concentration ratio of 0.5: 0.3 (bottom). The first column displays the substrate only, without any polyamide film for

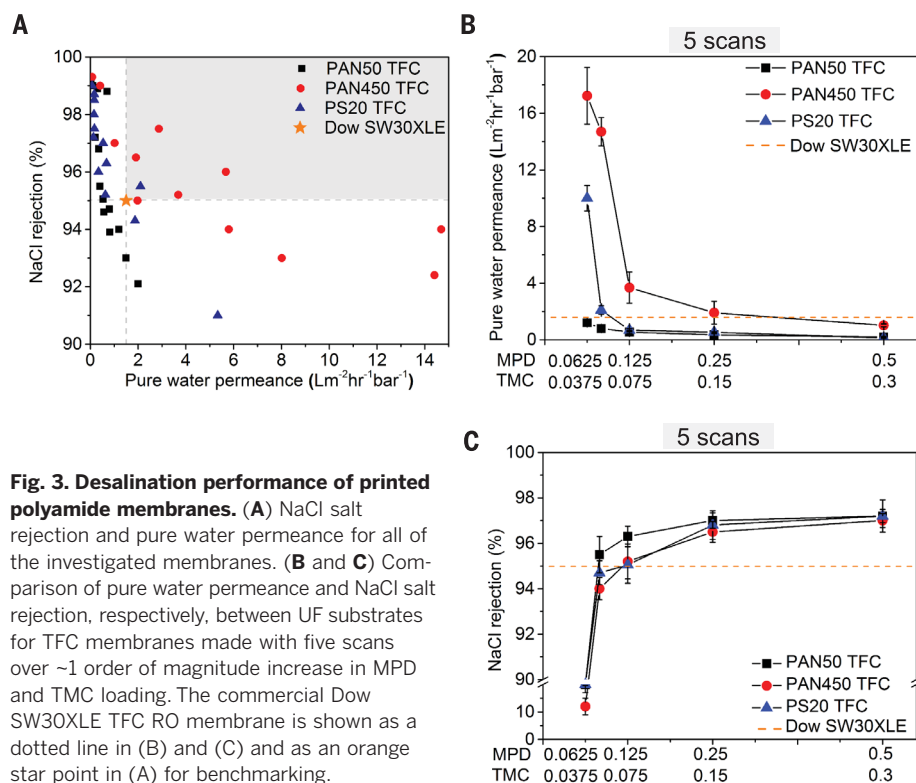
comparison. The inset numbers indicate either the concentration ratio or the number of scans. (C) Graph showing RMS surface roughness of the TFC membranes by using three different UF membranes as substrates for a series of MPD:TMC concentration ratios. The first points in the graph represent the roughness of the substrate only. (D) The surface roughness increases with the number of scans for three different MPD:TMC concentration ratios for PS20 TFC membranes. The commercial Dow SW30XLE TFC RO membrane is shown as a dotted line in (C) and (D) for benchmarking.

on monomer concentration. Higher concentrations of monomers form thicker (Fig. 3A) and less permeable films (Fig. 3B) while improving salt rejection (Fig. 3C). The efficacy of the TMC membranes can also be considered by redefining such data in terms of permselectivity, provided in fig. S10, where again these membranes similarly outperform conventional membranes.

The substrate selection has a noticeable effect on permeance. This is attributed to pore size and spacing on the substrate. The most permeable substrate (fig. S2C, PAN 450) exhibits the largest pores that are also closest together. This means that water diffusing through the film has less distance to travel to desorb through an open pore into the porous support, resulting in higher

permeance (6, 29, 30). These higher permeance values enabled our best performing membranes to match the upper-bound limit of the selectivity-permeability tradeoff relationship as described in (14) (fig. S11). Furthermore, there was no substrate effect on rejection as expected (Fig. 3C) because rejection is primarily a function of the selective film chemistry and structure. These film





**Fig. 3. Desalination performance of printed polyamide membranes.** (A) NaCl salt rejection and pure water permeance for all of the investigated membranes. (B and C) Comparison of pure water permeance and NaCl salt rejection, respectively, between UF substrates for TFC membranes made with five scans over  $\sim 1$  order of magnitude increase in MPD and TMC loading. The commercial Dow SW30XLE TFC RO membrane is shown as a dotted line in (B) and (C) and as an orange star point in (A) for benchmarking.

features are indistinguishable when deposited onto the three substrates.

Further tuning of desalination performance is done by changing the number of scans and hence polyamide thickness (fig. S12). Some of the thinnest membranes exhibited very high permeance, although the highest of these has correspondingly low salt rejection ( $\sim 10\%$ ). The TFC membranes made with five scans and an MPD:TMC ratio of 0.083:0.05 on the PAN 450 UF membrane exhibited a reasonable salt rejection of 94%, with a permeance of  $\sim 14.7$  liter  $m^{-2}$  hour $^{-1}$  bar $^{-1}$  (LMH bar $^{-1}$ ). This membrane also exhibited an RMS roughness only 2.3 nm higher than the substrate RMS roughness of 11.7 nm. This is less than one-sixth that of the SW30XLE membrane. Rejections as high as 95% were achieved on the same substrate for a MPD:TMC ratio of 0.125:0.075, with a RMS roughness only  $\sim 4.3$  nm greater than that of the substrate and a water permeance of 3.68 LMH bar $^{-1}$ . Increasing the number of scans to 10 yielded a salt rejection of 97.5% while still maintaining a water permeance of 2.87 LMH bar $^{-1}$  and a RMS roughness of less than 20 nm.

This additive approach to making TFC membranes has resulted in membranes with tunable

thickness and roughness while still retaining the selectivity expected of reverse osmosis membranes. These membranes have an intrinsic smoothness not seen in other TFC membranes today, can be tailored to thicknesses as low as 15 nm with as little as 4-nm resolution in thickness control, and can be formed on substrates without preparation. Furthermore, by decoupling the polyamide formation from the substrate properties, we have enabled the formation of TFCs on unconventional substrates and allowed for film characterization that would be impossible with polyamide films formed through conventional interfacial polymerization. The adaptation of this approach to other monomers or even simple polymers dissolved in solvents might enable the development of other TFC membranes for use in other separations.

#### REFERENCES AND NOTES

- P. W. Morgan, S. L. Kwolek, *J. Polym. Sci. Polym. Phys. Ed.* **XL**, 299–327 (1959).
- J. E. Cadotte, R. J. Petersen, R. E. Larson, E. E. Erickson, *Desalination* **32**, 25–31 (1980).
- J. E. Cadotte, Interfacially synthesized reverse osmosis membrane, U.S. patent 4,277,344 (1981).
- M. Elimelech, W. A. Phillip, *Science* **333**, 712–717 (2011).

- A. D. Khawaji, I. K. Kutubkhanah, J. M. Wie, *Desalination* **221**, 47–69 (2008).
- A. K. Ghosh, E. M. V. Hoek, *J. Membr. Sci.* **336**, 140–148 (2009).
- V. Freger, *Langmuir* **19**, 4791–4797 (2003).
- V. Freger, *Environ. Sci. Technol.* **38**, 3168–3175 (2004).
- J. E. Gu et al., *Adv. Mater.* **25**, 4778–4782 (2013).
- S. Karan, Z. Jiang, A. G. Livingston, *Science* **348**, 1347–1351 (2015).
- W. Choi et al., *J. Membr. Sci.* **527**, 121–128 (2017).
- X. Song, S. Qi, C. Y. Tang, C. Gao, *J. Membr. Sci.* **540**, 10–18 (2017).
- S.-J. Park et al., *J. Membr. Sci.* **526**, 52–59 (2017).
- T. Tsuru et al., *J. Membr. Sci.* **446**, 504–512 (2013).
- X.-H. Ma et al., *Environ. Sci. Technol. Lett.* **5**, 117–122 (2018).
- X.-H. Ma et al., *Environ. Sci. Technol. Lett.* **5**, 123–130 (2018).
- A. E. Seaver, C. J. Eckhardt, Electro spray coating process, U.S. patent 4,748,043 (1988).
- J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, C. M. Whitehouse, *Science* **246**, 64–71 (1989).
- J. B. Fenn, M. Mann, C. Meng, S. Wong, C. M. Whitehouse, *Mass Spectrom. Rev.* **9**, 37–70 (1990).
- R. Saf et al., *Nat. Mater.* **3**, 323–329 (2004).
- I. B. Rietveld, K. Kobayashi, H. Yamada, K. Matsushige, *Soft Matter* **5**, 593–598 (2009).
- J. Sakata, M. Mochizuki, *Thin Solid Films* **195**, 175–184 (1991).
- K. Morota et al., *J. Colloid Interface Sci.* **279**, 484–492 (2004).
- A. Jaworek, *Powder Technol.* **176**, 18–35 (2007).
- J.-U. Park et al., *Nat. Mater.* **6**, 782–789 (2007).
- A. Jaworek, A. T. Sobczyk, *J. Electrostat.* **66**, 197–219 (2008).
- Y. K. Hwang, U. Jeong, E. C. Cho, *Langmuir* **24**, 2446–2451 (2008).
- C. Y. Tang, Y. N. Kwon, J. O. Leckie, *J. Membr. Sci.* **287**, 146–156 (2007).
- J. Mulder, *Basic Principles of Membrane Technology* (Springer Netherlands, ed. 2, 1996).
- H. B. Park, J. Kamcev, L. M. Robeson, M. Elimelech, B. D. Freeman, *Science* **356**, eaab0530 (2017).

#### ACKNOWLEDGMENTS

The authors acknowledge M. Abril and X. Sun of Biosciences Electron Microscopy Facility at the University of Connecticut for performing TEM analysis. The SEM studies were performed by using the facilities in the University of Connecticut/Thermo Fisher Scientific Center for Advanced Microscopy and Materials Analysis (CAMMA). **Funding:** This work was supported by a University of Connecticut graduate teaching assistantship, General Electric Graduate Fellowship for Innovation, NSF DMR:MRI award 1726862, U.S. EPA grant RD834872, and the University of Connecticut Academic Plan funding program. **Author contributions:** All of the system design work, method development, model development, membrane fabrication, characterization, and testing were carried out by M.R.C. Experimental plan and theory were developed by M.R.C. and J.R.M. J.S. and B.D.H. provided AFM instruments, advice, training, experimental design assistance, and data analysis, and M.R.C. performed the experiments. **Competing interests:** M.R.C. and J.R.M. are inventors on a provisional patent application (U.S. 62/538503) submitted by University of Connecticut. **Data and materials availability:** All data needed for the conclusions made in this work are reported in the main text or the supplementary materials.

#### SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6403/682/suppl/DC1  
Materials and Methods  
Figs. S1 to S12  
Tables S1 to S4  
References (31–46)

28 October 2017; accepted 8 June 2018  
10.1126/science.aar2122

## PHOTOLUMINESCENCE

# Origin of the bright photoluminescence of few-atom silver clusters confined in LTA zeolites

Didier Grandjean<sup>1\*</sup>, Eduardo Coutiño-Gonzalez<sup>2</sup>, Ngo Tuan Cuong<sup>3,4</sup>, Eduard Fron<sup>2</sup>, Wouter Baekelant<sup>2</sup>, Saleh Aghakhani<sup>1</sup>, Philomena Schlexer<sup>5</sup>, Francesco D'Acapito<sup>6</sup>, Dipanjan Banerjee<sup>7</sup>, Maarten B. J. Roelfaers<sup>8\*</sup>, Minh Tho Nguyen<sup>4</sup>, Johan Hofkens<sup>2</sup>, Peter Lievens<sup>1\*</sup>

Silver (Ag) clusters confined in matrices possess remarkable luminescence properties, but little is known about their structural and electronic properties. We characterized the bright green luminescence of Ag clusters confined in partially exchanged Ag–Linde Type A (LTA) zeolites by means of a combination of x-ray excited optical luminescence-extended x-ray absorption fine structure, time-dependent-density functional theory calculations, and time-resolved spectroscopy. A mixture of tetrahedral  $\text{Ag}_4(\text{H}_2\text{O})_x^{2+}$  ( $x = 2$  and  $x = 4$ ) clusters occupies the center of a fraction of the sodalite cages. Their optical properties originate from a confined two-electron superatom quantum system with hybridized Ag and water O orbitals delocalized over the cluster. Upon excitation, one electron of the s-type highest occupied molecular orbital is promoted to the p-type lowest unoccupied molecular orbitals and relaxes through enhanced intersystem crossing into long-lived triplet states.

Few-atom luminescent silver clusters (AgCLs) (1) stabilized through organic (such as peptides, proteins, polymers, and DNA) (2–6) or inorganic (such as glasses and zeolites) (7–9) templates have emerged as promising candidates for a broad range of applications in lighting, imaging, sensing, and therapeutics (4). Compared with conventional quantum dots, few-atom AgCLs combine an ultrasmall size with excellent size-dependent photoluminescence (PL) spanning the ultraviolet to near-infrared spectrum. Strong quantum confinement of Ag valence electrons appears to break up the continuous density of states into discrete energy levels and confer molecular-like properties to the AgCLs. Nevertheless, the lack of a detailed understanding of the fundamental photophysical mechanisms underlying their emissions is hampering the rational design of AgCLs with improved and tailored optical properties. Atomic structures for AgCLs have not been determined unambiguously because of their vast distribution of size, environment, and tem-

plate interactions, as well as the presence of a large fraction of nonluminescent Ag species.

The AgCLs that self-assemble in the cavities of the rigid aluminosilicate crystalline framework of zeolites have the most homogeneous and efficient emissions. The PL of AgCLs confined in thermally activated Ag-loaded zeolite structures of faujasite (FAU) and Linde Type A (LTA) topologies features tunable absorption and emission, large Stokes shifts, and exceptionally high external luminescence quantum efficiencies reaching unity (8, 10). However, the structure of these AgCLs has not been fully elucidated yet because of the complexity of Ag-zeolite host-guest interactions and the sensitivity of Ag-zeolite composites toward radiation (electrons and photons) used in structural characterization techniques (11). We present a detailed investigation of the structural and electronic properties of partially Ag-exchanged  $\text{Ag}_3\text{K}_9$ -LTA zeolite by use of three complementary techniques. This system was selected for its green PL featuring an excellent external luminescence quantum yield of 23% among the Ag-LTA zeolites, its good stability toward x-ray irradiation (11), and its simpler crystallographic structure than that of FAU. With x-ray excited optical luminescence (XEOL), the x-ray absorption fine structure (EXAFS) signal is detected exclusively from the Ag atom fraction involved in the PL process at the Ag K-edge, thus selectively determining the structure of the emitting Ag species (12). This univocal assignment could not be made in earlier work (10); hence, approach may provide more detailed understanding of luminescent properties for a variety of few-atom Ag clusters. The structures obtained experimentally were confirmed computationally with geometry optimizations by using density functional theory (DFT) methods, while time-dependent DFT (TD-DFT) was applied

to determine the electronic transitions responsible for the absorption and emission spectra of the stable isomers. Last, to confirm the electronic structure of the theoretically modeled AgCLs, we identified the relevant decay modes and time scales involved in the absorption and luminescence processes of  $\text{Ag}_3\text{K}_9$ -LTA using a combination of femto- to millisecond time-resolved spectroscopies.

The large number of structural characterizations of AgCLs stabilized in LTA zeolites, often performed by means of x-ray diffraction and electron spin resonance (13–16), have led to numerous incomplete and often contradictory structural models.  $\text{Ag}_{3-4}$  clusters were tentatively related to partially Ag-exchanged LTA zeolites, whereas  $\text{Ag}_6$  clusters were associated to fully exchanged samples. High-resolution transmission electron microscopy (HRTEM) revealed octahedral  $\text{Ag}_6$  clusters in the sodalite cages of fully exchanged Ag-LTA zeolites (17), but no evidence linking this structure to the PL was given. Similar analysis of partially exchanged Ag-LTA zeolites remained unsuccessful because of the strong influence of electron-beam irradiation (17, 18).

EXAFS (19) provides information on cluster size and atom bonding both for the emissive and nonemissive clusters (20). By contrast, XEOL exclusively detects the XAFS signal from the atoms constituting the emissive species (12, 21, 22). Also, although x-ray irradiation can affect the structure of few-atom clusters (11), XEOL would monitor any beam degradation effect (supplementary materials). The three-dimensional (3D) structures of the AgCLs were determined by combining the fitting results of the XEOL and transmission-detected EXAFS collected simultaneously (table S1).

We primarily analyzed the XEOL-detected EXAFS of  $\text{Ag}_3\text{K}_9$ -LTA. The  $\chi(k)$   $k^3$ -weighted EXAFS data and the corresponding phase-corrected Fourier transform (FT) best fits are shown in Fig. 1, A and B.

The first and second peaks of the FT were fitted with, respectively, two oxygen (O) atoms at 2.36 Å (N1) and three Ag atoms at 2.82 Å (N2). Additionally, the fit was completed with two longer shells consisting of 0.4 K at 3.05 Å (N4) and 1.1 Ag at 3.3 Å (N5). The Ag atoms coordinated to three other Ag atoms ( $\text{Ag}_c$ ) form tetrahedral-like  $\text{Ag}_4$  clusters located inside the sodalite cage, as was shown with TEM (17). The  $\text{Ag}_c$  atoms are each coordinated to two O atoms likely corresponding to extra-framework water molecules because of the short Ag–O distances and the fact that these O atoms are removed with a concomitant loss of the sample PL upon dehydration of the material (supplementary materials). The absence of contributions from the sodalite atoms [O, silicon (Si), and aluminum (Al)] embedding AgCLs in the EXAFS signal is discussed in the supplementary materials.

The analysis of the XEOL-detected signal shows that the species at the origin of the bright-green PL observed in  $\text{Ag}_3\text{K}_9$ -LTA are  $\text{Ag}_4$  clusters with short Ag–Ag distances of 2.82 Å, in which each Ag atom is bound to two water molecules at 2.36 Å. They are further surrounded by isolated K

<sup>1</sup>Laboratory of Solid State Physics and Magnetism, KU Leuven, Celestijnenlaan 200D, B-3001 Leuven, Belgium.

<sup>2</sup>Molecular Visualization and Photonics, KU Leuven, Celestijnenlaan 200F, B-3001 Leuven, Belgium. <sup>3</sup>Faculty of Chemistry, Hanoi National University of Education, 136 Xuan Thuy, Hanoi, Vietnam. <sup>4</sup>Department of Chemistry, KU Leuven, Celestijnenlaan 200F, B-3001 Leuven, Belgium. <sup>5</sup>Dipartimento di Scienza dei Materiali, Università di Milano-Bicocca, via Cozzi, 55, 20125 Milano, Italy. <sup>6</sup>Consiglio Nazionale delle Ricerche-Istituto Officina dei Materiali-Operative Group in Grenoble (CNR-IOM-OGG), European Synchrotron Radiation Facility (ESRF) LISA Collaborating Research Group, Grenoble, France. <sup>7</sup>Dutch-Belgian beamline, ESRF-The European Synchrotron, CS 40220, 38043 Grenoble, France. <sup>8</sup>Centre for Surface Chemistry and Catalysis, KU Leuven, Celestijnenlaan 200F, B-3001 Leuven, Belgium.

\*Corresponding author. Email: didier.grandjean@kuleuven.be (D.G.); maarten.roelfaers@kuleuven.be (M.B.J.R.); peter.lievens@kuleuven.be (P.L.)

and Ag cations likely positioned in the single six-membered rings (S6Rs) of the same sodalite cage. A twofold coordination of Ag<sub>C</sub> atoms, however, does not correspond to an integer number of water molecules but rather to two different stoichiometries of  $x = 2$  and  $x = 4$ , corresponding to a water coordination per Ag<sub>C</sub> of 1.5 and 3, respectively. This analysis suggests the presence of a mixture of Ag<sub>4</sub>(H<sub>2</sub>O)<sub>4</sub> and Ag<sub>4</sub>(H<sub>2</sub>O)<sub>2</sub> with a ~34/66 ratio ( $3 \times 0.34 + 1.5 \times 0.66 = 2$ ). Attempts to use other models for analyzing the XEOL-detected signal, including the structure used for the analysis of the transmission-detected EXAFS presented below, were either not successful or led to unrealistic fitting parameters (supplementary materials).

We also analyzed the transmission-detected EXAFS collected simultaneously with the XEOL-detected data. The  $\chi(k)$   $k^3$ -weighted EXAFS data and the phase-corrected FT best fits of heat-treated Ag<sub>3</sub>K<sub>9</sub>-LTA are shown in Fig. 1, C and D. The distinct profiles compared with those collected with XEOL detection indicate that two different average local environments of Ag atoms were measured simultaneously, which is consistent with the x-ray absorption near-edge structure analysis (supplementary materials).

The first peak in the FT was fit with 2.5 O at 2.34 Å (N1) corresponding to the combination of the framework O (O<sub>F</sub>) from the S6Rs rings and the H<sub>2</sub>O ligands, as shown in the XEOL analysis. The second peak in the FT is a multiplex composed of 2.6 Si/Al atoms (N2) at 3.26 to 3.30 Å corresponding to a fraction of non-luminescent Ag cations located near the center of the S6Rs (Ag<sub>R</sub>) (fig. S13), not detected in XEOL-EXAFS. The second peak in the FT analysis also contained a weaker Ag-Ag contribution (N3) of 1.7 Å at unusually short distances of 2.70 Å. This feature corresponds to the remaining part of the silver atoms Ag<sub>C</sub> (~57%) that are coordinated to ~3 (1.7/0.57) silver neighbors and are forming Ag<sub>4</sub> clusters inside the sodalite cage. The 4% discrepancy between the Ag<sub>C</sub>-Ag<sub>C</sub> (2.70 to 2.82 Å) distances found by the two detection approaches suggests that XEOL measured preferentially the excited state structure of the cluster.

(supplementary materials). Ag<sub>C</sub> in Ag<sub>4</sub> clusters are coordinated to 2.1 O (1.2/0.57) from water molecules. Additionally, four shells (N<sub>4</sub> to N<sub>7</sub>) consisting of K and Ag were detected at liquid-nitrogen temperatures between 2.97 and 4.49 Å corresponding to Ag<sub>C</sub>-K or Ag<sub>C</sub>-Ag<sub>R</sub> distances from basal Ag<sub>C</sub> in Ag<sub>4</sub> tetrahedra associated with the absence or presence, respectively, of a water molecule sandwiched between the two atoms (fig. S13). These shells are complemented by two long-distance contributions (N<sub>8</sub> and N<sub>9</sub>) corresponding to Ag<sub>C</sub>-Ag<sub>R</sub> from apical Ag<sub>C</sub> in Ag<sub>4</sub> tetrahedra and Ag<sub>R</sub>-Ag<sub>R</sub> (fig. S7a) detected at 5.24 and 6.13 Å at LN. The distinct combination of distances of N<sub>4</sub> to N<sub>9</sub> shells closely fit the Ag-LTA sodalite crystallographic model (fig. S7B), fully supporting the AgCL local structures proposed.

The EXAFS investigation shows that the emitters in Ag<sub>3</sub>K<sub>9</sub>-LTA consist of ~40% of Ag<sub>4</sub>(H<sub>2</sub>O)<sub>4</sub> and 60% of Ag<sub>4</sub>(H<sub>2</sub>O)<sub>2</sub> tetrahedra-like clusters located at the center of the sodalite cage. These structures are presented in Fig. 1, E to J. The clusters are coordinated at their faces by two or four water molecules located near the center of the S6Rs and sandwiched between three Ag<sub>C</sub> and one Ag<sub>R</sub> (or K cation). AgCLs consisting of 57% of the total number of exchanged Ag atoms are mostly surrounded by the remaining 43% isolated Ag<sub>R</sub> cations plus some additional K cations in the S6Rs. This indicates that Ag cations in partially exchanged Ag<sub>3</sub>K<sub>9</sub>-LTA concentrate (six to seven Ag atoms instead of three expected from the Ag stoichiometry) in a limited fraction of the sodalite cages (~45%). (23).

We used a combination of DFT and TD-DFT to model AgCLs and probed their charges with the natural bonding orbital approach. Two stable isomers—[Ag<sub>4</sub>(H<sub>2</sub>O)<sub>x</sub>(Si<sub>24</sub>H<sub>24</sub>O<sub>36</sub>)],  $x = 2$  and  $x = 4$ , showing the best agreement between calculated and measured structures and absorption spectra—were obtained when applying a +2 charge preferentially localized on the Ag<sub>4</sub>CLs but extending toward the cluster surrounding (tables S4 and S5). The doubly charged Ag<sub>4</sub>CLs exhibit a closed-shell electronic configuration, in which the Ag 4d shell is completely filled, and the two remaining valence 5s electrons are delocalized over the

cluster. Within a superatom model, two electrons associated to a metal cluster correspond to the smallest magic number with enhanced stability (24, 25), which is consistent with analogous theoretical work on Ag<sub>4</sub>CLs (26, 27).

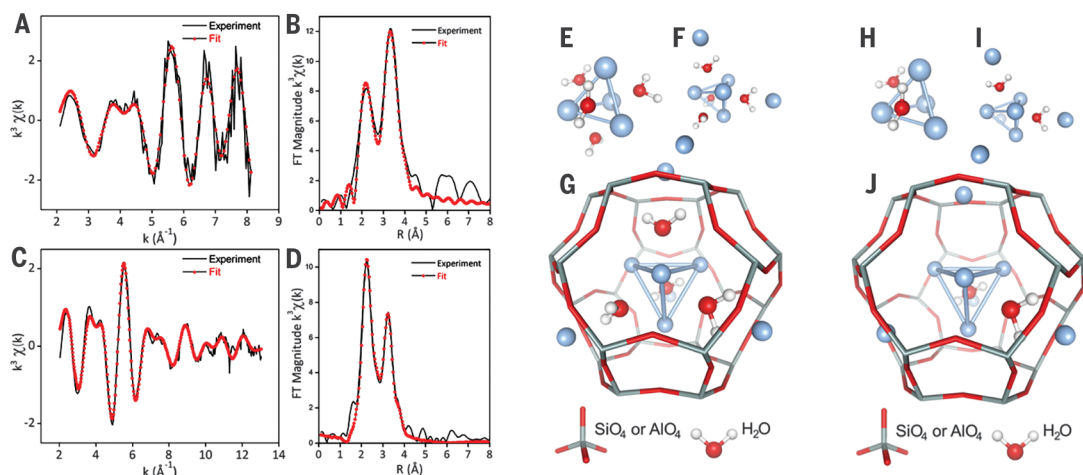
Both isomers [Ag<sub>4</sub>(H<sub>2</sub>O)<sub>x</sub>,  $x = 2$  and 4] consist of pseudotetrahedral Ag<sub>4</sub> clusters located, unlike the experimental structures, in an off-centered position in the sodalite cage, with one or two Ag atoms coordinated directly to O<sub>F</sub> (supplementary materials). In Ag<sub>4</sub>(H<sub>2</sub>O)<sub>4</sub>, each Ag atom is coordinated on average to two O atoms (water plus O<sub>F</sub>) with an average Ag-Ag bond distance of ~2.79 and 2.92 Å in the ground and excited state, respectively (fig. S14). In Ag<sub>4</sub>(H<sub>2</sub>O)<sub>2</sub>, an average Ag-Ag distance of 2.87 Å in the ground state and a mean O coordination close to 1.3 were obtained (fig. S15). These calculated structures featuring a 5% increase in the Ag<sub>C</sub>-Ag<sub>C</sub> distances from the ground to the excited states are in good agreement with the experimental results (supplementary materials).

The calculated frontier orbitals for both Ag<sub>4</sub>(H<sub>2</sub>O)<sub>4</sub><sup>2+</sup> and Ag<sub>4</sub>(H<sub>2</sub>O)<sub>2</sub><sup>2+</sup> isomers (Fig. 2A and fig. S16, respectively) are composed of a contribution of ~50% from Ag 5s atomic orbitals and of up to 25% from the O states of the surrounding O<sub>F</sub> and H<sub>2</sub>O, as shown in the density of states curves (fig. S18). A doubly occupied HOMO of totally symmetric s-type orbital forms the ground state (<sup>1</sup>S<sub>0</sub>) and two sets of three singlet (<sup>1</sup>P) and three triplet (<sup>3</sup>P) LUMOs consisting of one-node p-type orbitals form the expected lowest-lying cluster orbitals for a cluster system with two skeleton electrons. Both states in each isomer have similar atomic orbital compositions (fig. S18), suggesting that the excitations are mainly localized on the cluster.

The key role of water ligands in AgCLs' electronic properties is highlighted in the energy level diagram (Fig. 2B and fig. S17) of Ag<sub>4</sub>(H<sub>2</sub>O)<sub>4</sub><sup>2+</sup> and Ag<sub>4</sub>(H<sub>2</sub>O)<sub>2</sub><sup>2+</sup> along with water-free Ag<sub>4</sub><sup>2+</sup> clusters as reference. In water-free unperturbed Ag<sub>4</sub><sup>2+</sup>, the ground state is the HOMO <sup>1</sup>S<sub>0</sub>, and the LUMOs consist of threefold degenerated singlet <sup>1</sup>P and triplet <sup>3</sup>P excited states. Upon coordination with water, the ligand field lifts the degeneracy of the

**Fig. 1. Ag K-edge XEOL and transmission-detected EXAFS and FTs of heat-treated Ag<sub>3</sub>K<sub>9</sub>-LTA and derived structures.**

(A) XEOL-detected and (C) transmission-detected  $k^3$ -weighted Ag K-edge EXAFS with the (B) phase-corrected XEOL-detected FT and (D) transmission-detected FT best fits. (E to J) Structures of (E) Ag<sub>4</sub>(H<sub>2</sub>O)<sub>4</sub> and (H) Ag<sub>4</sub>(H<sub>2</sub>O)<sub>2</sub>, including [(F) and (I)] Ag<sub>R</sub> cations and [(G) and (J)] embedded in the sodalite cage (~0.66 nm free diameter).



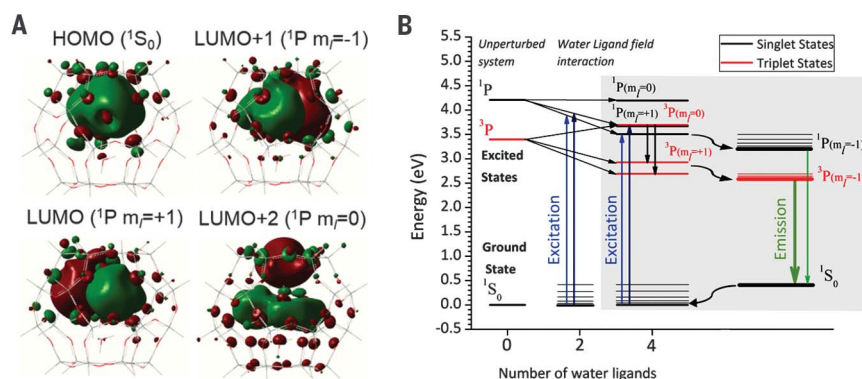


LUMOs into six excited states: three singlet  $^1P$  [total spin quantum number ( $S$ ) = 0; total orbital angular momentum quantum number ( $L$ ) = 1; magnetic quantum number ( $m_l$ ) = -1, +1, or 0] and three triplet  $^3P$  ( $S$  = 1;  $L$  = 1;  $m_l$  = -1, +1, or 0) states. The absorption occurs from the ground state corresponding to two electrons of opposite spins on a HOMO  $s$ -like orbital ( $^1S_0$ ) to the singlet excited states  $^1P$  (3.5 and 3.7 eV) that correspond to one electron on a  $s$ -like orbital and one electron on LUMO, LUMO+1, or LUMO+2  $p$ -like orbitals, with the two electrons having opposite spins.

These transitions feature large oscillator strengths  $f$  because they are allowed by spin and angular momentum selection rules. The strong overlap of the high-energy triplet  $^3P$  ( $S$  = 1;  $L$  = 1;  $m_l$  = 0) state with the  $^1P$  singlet states ensures, in combination with large spin-orbit coupling expected for Ag, an enhanced intersystem crossing. Upon light excitation, a fraction of  $^1P$  singlet states population is transferred to the high-energy triplet state that finally decays into the low-lying  $^3P$  ( $S$  = 1,  $L$  = 1,  $m_l$  = -1 or +1) triplet state. As shown in Fig. 2B, the intensity of the ligand-field splitting of the  $p$ -like LUMOs is directly proportional to the number of water molecules coordinating AgCLs, decreasing the band gap when going from  $Ag_4(H_2O)_2^{2+}$  to  $Ag_4(H_2O)_4^{2+}$  isomers. The bright-green emission of  $Ag_4(H_2O)_4^{2+}$  then occurs from the lowest-lying  $^3P$  triplet excited state  $^3P$  ( $S$  = 1,  $L$  = 1,  $m_l$  = -1) to the ground state  $^1S_0$ .

The modeled absorption spectra for the  $Ag_4(H_2O)_4^{2+}$  and  $Ag_4(H_2O)_2^{2+}$  isomers (Fig. 3) are in excellent agreement with the steady-state optical experimental data. Calculated absorption peaks found at 343 and 320 nm in  $Ag_4(H_2O)_4^{2+}$  and  $Ag_4(H_2O)_2^{2+}$ , respectively, closely match the experimental excitation peaks at 340 nm (3.64 eV) and 310 nm (4.00 eV). These results also confirm the assumption that the 37/63 intensity ratio of the two main emission peaks is directly related to the 34/66 fraction ratio of  $Ag_4(H_2O)_4$  and  $Ag_4(H_2O)_2$  present in  $Ag_3K_9$ -LTA composites. Last, the green PL energy obtained experimentally at 550 nm (2.25 eV) fits the transition energy of 556 nm (2.23 eV) of the modeled transition from the relaxed lowest-lying  $^3P$  triplet excited state to the  $^1S_0$  ground state (Fig. 2B).

To determine the decay time scale and energies of  $Ag_3K_9$ -LTA modeled optical transitions and to verify the triplet nature of its bright green PL, we performed a global analysis of the decays obtained with femtosecond fluorescence up-conversion, time-correlated single-photon counting (TCSPC) and nanosecond luminescence time-resolved spectroscopies. Decays obtained with femtosecond fluorescence up-conversion in the range of 410 to 570 nm reveal three time components of 0.5 and 2.6 ps related to relaxation processes and >50 ps attributed to the  $^1P$ -to- $^1S_0$  transition (Fig. 4, A and B; fig. S24, decay traces; and table S7). The amplitude-to-wavelength dependence (AWD) of the first two ultrafast components (maxima centered at 510 and 530 nm) are slightly blue-shifted relative to the maximum

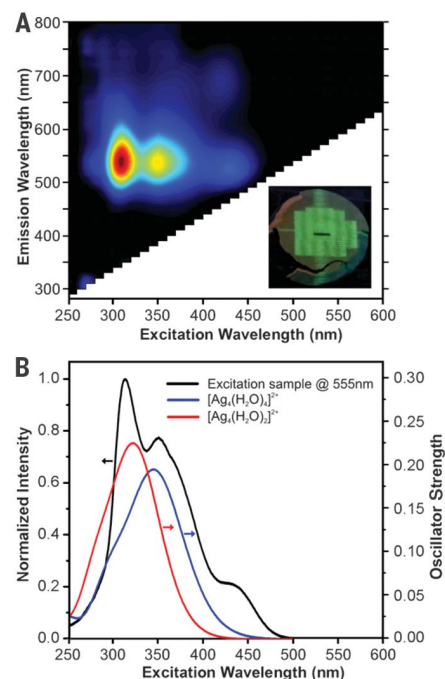


**Fig. 2. Frontier orbitals of  $[Ag_4(H_2O)_4(Si_{24}H_{24}O_{36})]^{2+}$  and  $Ag_4(H_2O)_4^{2+}$  clusters in  $Ag_3K_9$ -LTA.** (A) Frontier orbitals consist of one single symmetric  $s$ -type HOMO ( $^1S_0$ ) and three singlet one-node  $p$ -type  $^1P$  ( $m_l$  = -1, +1, or 0) LUMOs ( $p_x$ ,  $p_y$ ,  $p_z$ ) delocalized over all the Ag and O atoms of the cluster. Atom colors are Si, gray; O, red; Ag, blue; hydrogen, white. (B) Energy level diagram showing the ground-state  $^1S_0$  and the excited states  $^1P$  and  $^3P$  of water-free unperturbed  $Ag_4^{2+}$  clusters and the ground-state  $^1S_0$  and the six singlet  $^1P$  and triplet  $^3P$  excited states of  $Ag_4(H_2O)_2^{2+}$  and  $Ag_4(H_2O)_4^{2+}$  perturbed by means of water ligand field interaction. Blue arrows represent the allowed transitions, and the green arrows represent the luminescent transitions between the relaxed states.

of the steady-state emission spectrum (550 nm), suggesting the rapid depopulation of  $^1P$  ( $S$  = 0;  $L$  = 1;  $m_l$  = -1, +1) singlet excited states. These ultrafast components attributed to nonfluorescent short-lived intermediate species formed after the relaxation of  $^1P$  Franck-Condon states match closely the energy of the  $^1P$  ( $S$  = 0,  $L$  = 1,  $m_l$  = -1)-to- $^1S_0$  transition predicted with TD-DFT (Fig. 2B). These intermediate states rapidly convert into fluorescent  $^3P$  triplet states lying at similar energies via intersystem crossing (Fig. 4, B and D).

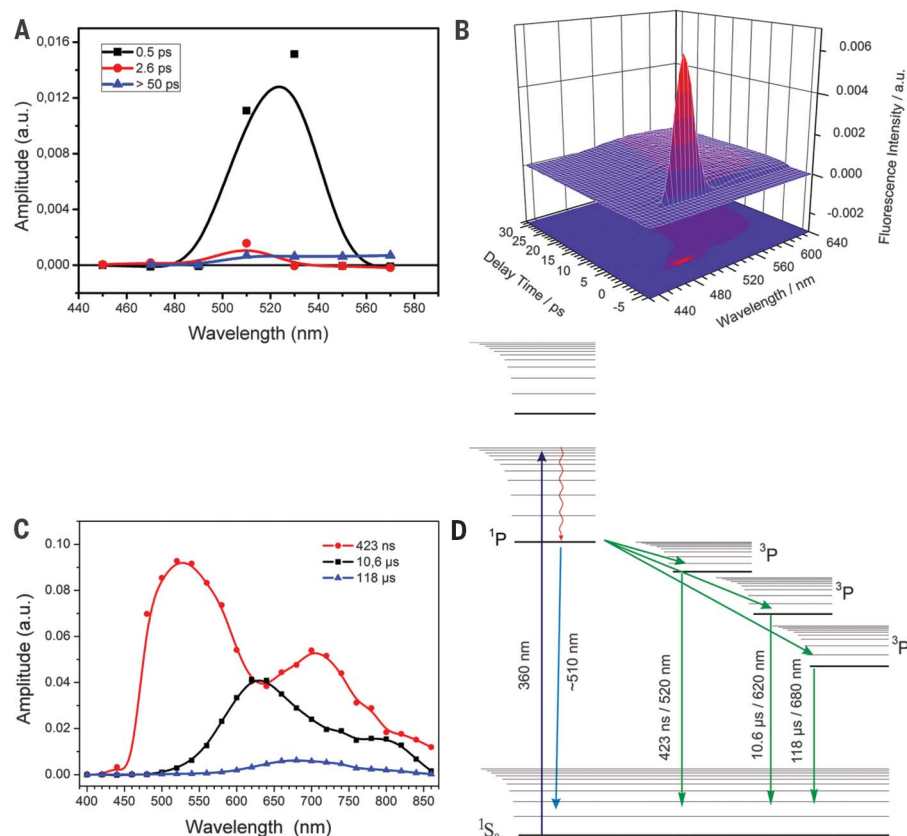
This model is confirmed through the analysis of the PL decays in the micro- to millisecond range that reveal three time constants of 423 ns, 10.6  $\mu$ s, and 116  $\mu$ s. On the basis of the AWD presented in Fig. 4C, the 423-ns component is attributed to an excited state with an intense emission peaking at 520 to 540 nm (2.38 to 2.30 eV), with estimated radiative and nonradiative rates of  $5.43 \times 10^5$  s $^{-1}$  and  $18.2 \times 10^5$  s $^{-1}$ , respectively. The close resemblance of the AWD of this state with the stationary emission spectrum indicates that these long-lived species are at the origin of the bright-green emission observed in  $Ag_3K_9$ -LTA. This emission occurring from long-lived states characteristic for spin-forbidden transitions points toward their peculiar triplet nature. The other decay times of 10.6 and 118  $\mu$ s are attributed to two different triplet excited states, with weak emissions centered at 630 and 680 nm (1.97 and 1.82 eV), respectively (Fig. 2B; fig. S25, decay traces; and table S8), which are associated with the presence of residual amounts of emissive species such as AgCLs with different size and/or water coordination (10).

The triplet nature of  $Ag_3K_9$ -LTA bright-green emission is further corroborated by the remarkable enhancement of the PL accompanied by a dramatic increase of the decay time from 423 ns to 106  $\mu$ s for the 540-nm main emission at low temperature (77 K), whereas the faster components remained mostly unaffected (figs. S26 to S33).



**Fig. 3. Steady-state excitation-emission of  $Ag_3K_9$ -LTA.** (A) 2D excitation-emission plot. (Inset) The picture of an x-ray-irradiated sample under 366 nm illumination. (B) Excitation spectrum  $\lambda_{\text{detection}} = 555$  nm of as-prepared  $Ag_3K_9$ -LTA. Calculated  $^1S_0$  HOMO to  $^1P$  ( $S$  = 0;  $L$  = 1;  $m_l$  = -1, +1) LUMOs absorption spectra of  $x = 2$  and  $x = 4$  [ $Ag_4(H_2O)_x$ ] $^{2+}$  isomers showing a good agreement with experiments.

This result shows that excited-state kinetics, in which nonradiative decay channels are hindered at low temperature and characteristic of triplet-state emissions, is involved. Hence, given the close correspondence between the computational and



**Fig. 4. Time-resolved spectroscopy of  $\text{Ag}_3\text{K}_9\text{-LTA}$ .** (A and B) AWD 3D time-resolved fluorescence emission spectra in 50-ps time window obtained with (C) femtosecond fluorescence up-conversion, in 1-ms time window through nanosecond luminescence. (D) Schematic illustration of the main electronic states involved as a function of energy (on the vertical axis).

the photophysics results, the cluster-ligand interaction as included in the model of a two-electron doubly charged  $\text{AgCl}$ , possibly with a fraction of the charge located on the sodalite cage, explains the occurrence of long-lived bright luminescent states in  $\text{Ag}_3\text{K}_9\text{-LTA}$ .

By measuring exclusively the local structure of the emissive Ag clusters in partially exchanged LTA zeolites, XEOL-XAFS has allowed for the first time the unambiguous identification of their functional structures. DFT modeling based on these detailed structures showed that the double positively charged  $\text{Ag}_4(\text{H}_2\text{O})_4$  and  $\text{Ag}_4(\text{H}_2\text{O})_2$  clusters, in which water ligand molecules modulate the HOMO-LUMO gap, behave as confined two-electron helium or alkaline earth-like superatom quantum systems that mainly emit via their long-lived lowest-lying  $^3\text{P}$  triplet excited state, as confirmed with time-resolved optical spectroscopy. We anticipate that similar photophysical properties may also apply to luminescent AgCLs con-

finied in other inorganic and organic scaffolds. This is likely the case for AgCLs confined in fully exchanged LTA or in FAU zeolites that possess very similar structural and luminescent properties (10). This new understanding of the mechanism of AgCLs' bright luminescence and its expected dependence on the interaction with oxygen ligands, electron confinement, electrostatic interaction, and charge transfer to the surrounding silver atoms, which differ from the single Ag cations emission model proposed earlier (28–30), should lead to substantial improvements in the rational design of AgCLs optical properties.

#### REFERENCES AND NOTES

1. L. A. Peyser, A. E. Vinson, A. P. Bartko, R. M. Dickson, *Science* **291**, 103–106 (2001).
2. C. I. Richards et al., *J. Am. Chem. Soc.* **130**, 5038–5039 (2008).
3. T. Vosch et al., *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12616–12621 (2007).
4. Y. Yu, B. Y. L. Mok, X. J. Loh, Y. N. Tan, *Adv. Healthc. Mater.* **5**, 1844–1859 (2016).

5. I. Díez et al., *Nanoscale* **4**, 4434–4437 (2012).
6. J. Yu, S. A. Patel, R. M. Dickson, *Angew. Chem. Int. Ed.* **46**, 2028–2030 (2007).
7. A. Simo et al., *J. Am. Chem. Soc.* **134**, 18824–18833 (2012).
8. G. De Cremer et al., *J. Am. Chem. Soc.* **131**, 3049–3056 (2009).
9. A. S. Kuznetsov, V. K. Tikhomirov, M. V. Shestakov, V. V. Moshchakov, *Nanoscale* **5**, 10065–10075 (2013).
10. O. Fenwick et al., *Nat. Mater.* **15**, 1017–1022 (2016).
11. E. Coutino-Gonzalez et al., *Chem. Commun. (Camb.)* **50**, 1350–1352 (2014).
12. T. K. Sham, *Adv. Mater.* **26**, 7896–7901 (2014).
13. T. Sun, K. Seff, *Chem. Rev.* **94**, 857–870 (1994).
14. P. A. Jacobs, J. B. Uytterhoeven, H. K. Beyer, *J. Chem. Soc., Faraday Trans. 1* **75**, 56 (1979).
15. J. Michalik, L. Kevan, *J. Am. Chem. Soc.* **108**, 4247–4253 (1986).
16. P. J. Grobet, R. A. Schoonheydt, *Surf. Sci.* **156**, 893–898 (1985).
17. A. Mayoral, T. Carey, P. A. Anderson, A. Lubk, I. Diaz, *Angew. Chem. Int. Ed.* **50**, 11230–11233 (2011).
18. L. A. Bursill, J. M. Thomas, K. J. Rao, *Nature* **289**, 157–158 (1981).
19. A. Baldansuren, H. Dilger, R. A. Eichel, J. A. van Bokhoven, E. Roduner, *J. Phys. Chem. C* **113**, 19623–19632 (2009).
20. M. L. Neidig et al., *J. Am. Chem. Soc.* **133**, 11837–11839 (2011).
21. L. Liu, T. K. Sham, W. Han, C. Zhi, Y. Bando, *ACS Nano* **5**, 631–639 (2011).
22. T. K. Sham, R. A. Rosenberg, *ChemPhysChem* **8**, 2557–2567 (2007).
23. M. Meyer, C. Leiggener, G. Calzaferri, *ChemPhysChem* **6**, 1071–1080 (2005).
24. M. Walter et al., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9157–9162 (2008).
25. T. Yumura, M. Kumondai, Y. Kuroda, T. Wakasugi, H. Kobayashi, *RSC Advances* **7**, 4950–4959 (2017).
26. A. Kulesza et al., *Angew. Chem. Int. Ed.* **50**, 878–881 (2011).
27. P. Schlexer, A. Ruiz Puigdollers, G. Pacchioni, *Phys. Chem. Chem. Phys.* **17**, 22342–22360 (2015).
28. H. Lin, K. Imakita, M. Fujii, *Appl. Phys. Lett.* **105**, 211903 (2014).
29. H. Lin et al., *Nanoscale* **7**, 15665–15671 (2015).
30. R. Seifert, R. Rytz, G. Calzaferri, *J. Phys. Chem. A* **104**, 7473–7483 (2000).

#### ACKNOWLEDGMENTS

We thank B. Dieu for the graphical material. **Funding:** We acknowledge the IAP-7 (Belspo), the EU (FP7/2007-2013 and GA nos. 310651-SACS and 607417-Catsense), the Flemish government “Methusalem” (CASAS, Meth/08/04, CASAS2, and Meth/15/04), the SoPpM program, the FWO (G.0990.11, G.0197.11, G.0962.13, G.0B39.15, and ZWI15\_09 GOH6316N), and KU Leuven (GOA/14 and IDO/07/011). We thank the ESRF (CH-4207) and the staff of LISA-BM08 and DUBBLE-BM26A (26-01-865) beamlines. **Author contributions:** D.G. and P.L. conceived and directed the study. E.C.-G., W.B., and M.B.J.R. prepared the samples and performed the steady-state optical, elemental, and thermogravimetric analysis characterization. D.G., E.C.-G., F.D., D.B., W.B., and S.A. performed the XEOL-XAFS and Tr-XAFS measurements. D.G. and S.A. analyzed the XAFS data. N.T.C., P.S., and M.T.N. performed the DFT calculations. E.F. and J.H. designed, performed, and analyzed the time-resolved optical experiments. All authors contributed to discussions and interpretations of the combined experimental and theoretical results. D.G., P.L., E.F., and S.A. prepared the manuscript with contributions from all coauthors. All authors have approved the final version of the manuscript. **Competing interests:** None declared. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

#### SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6403/686/suppl/DC1  
Materials and Methods  
Supplementary Text  
Figs. S1 to S35  
Tables S1 to S8  
References (31–45)

13 November 2017; accepted 15 June 2018  
10.1126/science.aag1308

## SOLID-STATE PHYSICS

# Twistable electronics with dynamically rotatable heterostructures

Rebeca Ribeiro-Palau<sup>1,2\*,†</sup>, Changjian Zhang<sup>2,3\*</sup>, Kenji Watanabe<sup>4</sup>, Takashi Taniguchi<sup>4</sup>, James Hone<sup>2</sup>, Cory R. Dean<sup>1†</sup>

In heterostructures of two-dimensional materials, electronic properties can vary dramatically with relative interlayer angle. This effect makes it theoretically possible to realize a new class of twistable electronics in which properties can be manipulated on demand by means of rotation. We demonstrate a device architecture in which a layered heterostructure can be dynamically twisted *in situ*. We study graphene encapsulated by boron nitride, where, at small rotation angles, the device characteristics are dominated by coupling to a long-wavelength moiré superlattice. The ability to investigate arbitrary rotation angle in a single device reveals features of the optical, mechanical, and electronic response in this system not captured in static rotation studies. Our results establish the capability to fabricate twistable electronic devices with dynamically tunable properties.

**T**he weak van der Waals forces between the atomic planes in two-dimensional (2D) materials make it possible to fabricate devices with arbitrary rotational order. Devices could be designed in which electronic properties are controlled by varying the relative twist angle between layers (1). Several studies have established that in heterostructures assembled from 2D crystals, electron tunneling between layers varies strongly with rotation (2–8). In twisted bilayer graphene (two monolayers in direct contact but with an angle mismatch between the layers), several new phenomena have been predicted and observed, including topological valley transport (9–13) and superconductivity (14), as a consequence of angle-dependent interlayer coupling. Likewise, the formation of interlayer excitons in transition metal dichalcogenide heterostructures is highly sensitive to angle (15–17).

The effect of rotational alignment between conducting and insulating 2D layers can be equally important. An example is provided by graphene coupled to hexagonal boron nitride (BN). The closely matched lattice constants create a large moiré superlattice that develops near zero angle mismatch (18, 19). This superlattice substantially alters the graphene band structure, opening an energy gap at the charge neutrality point (CNP) and creating replica Dirac points at higher energies (20–22).

Several techniques have been developed to fabricate layered heterostructures with controlled rotation between the layers, including optical alignment of crystal edges (18, 20–22), rota-

tional alignment (23) during assembly, and self-alignment through thermal annealing (24, 25). However, in each case, an *a priori* understanding of the crystallographic orientation of each layer is required before assembly; motion between the layers during assembly makes it difficult to achieve precise angle control; and, most importantly, once assembled, the angle cannot be further modified. Here we present an experimental technique that provides on-demand control of the orientation between layers in a van der Waals heterostructure. In a BN/graphene/BN structure, we demonstrate *in situ* control over the length of the moiré potential and thus achieve dynamic tunability of the optical, mechanical, and electronic properties of the system.

Figure 1A shows a cartoon schematic of our device design. Using the mechanical assembly technique described in (26), graphene was placed on top of a large flake of BN, then etched into a Hall bar shape with an oxygen plasma. The graphene layer is intentionally misaligned to this BN, producing a short-wavelength moiré potential that does not substantially alter the intrinsic graphene band structure (27). Next, a preshaped BN structure was transferred on top of the graphene. Last, electrical contacts were patterned onto the exposed leads of the graphene (fig. S1). Because of the low mechanical friction between graphene and BN, we could freely rotate and translate this top BN layer by using an atomic force microscope (AFM). Pushing on one of the arms of the uppermost BN structure rotates this layer (Fig. 1, B to D) and changes its crystallographic orientation with respect to the graphene layer. The moiré superlattice wavelength,  $\lambda$ , generated between these crystals is given by

$$\lambda = \frac{(1 + \delta)a}{\sqrt{2(1 + \delta)[1 - \cos(\theta)] + \delta^2}} \quad (1)$$

where  $\delta = 0.017$  is the lattice mismatch between graphene and BN,  $a$  is the lattice constant of

graphene, and  $\theta$  is the rotational mismatch between the layers. By rotating the top BN layer, the wavelength of the resulting moiré superlattice can be dynamically varied.

Imaging with the same AFM provided a real-time measurement of the orientation of the rotatable BN layer relative to the reference frame of the AFM, which we label as the absolute angle  $\theta_A$  (for example, Fig. 1, B to D). More important is the relative angle,  $\theta$ , between the rotatable BN and the encapsulated graphene crystal lattices. To determine this angle, we identified and measured angle-dependent features in the optical Raman spectrum of the heterostructure.

Figure 1F shows a series of Raman spectra measured for  $\theta_A$  from 34° to 64°. The most striking variation in the Raman spectra was an increase in the full width at half maximum of the 2D peak (FWHM<sub>2D</sub>) (19, 28); a well-defined maximum occurred for this device at ~34.2° (Fig. 1G). On the basis of previous work (28), we interpret the peak position as corresponding to zero-angle alignment between the BN and graphene crystals. This peak provides a reference that we could use to determine  $\theta$  for any orientation of the BN layer.

Our results highlight the robustness of Raman spectroscopy as a tool to characterize the rotational order in these van der Waals heterostructures. Moreover, previous efforts to study effects of twist angle in this system required numerous samples with different fixed angles, whereas here we demonstrate a mapping of the angular dependence with better than 0.2° precision in a single tunable device. One notable disagreement with previous results (28) is an overall narrower 2D linewidth in our devices. In previous work, this reduced linewidth was proposed to be a consequence of a reduction of the in-plane strain in fully encapsulated structures, so that areas of full commensuration to the aligned BN disappear (19). However, our observation of the same linear trend and a reduced linewidth for the misaligned position, for which no commensurate regions are expected, suggests a different origin. We speculate that this linewidth reduction was caused by the change in the dielectric environment of the graphene caused by the presence of the second BN, as proposed in (29). Thus, our results reopen the question of the existence of commensurate states in encapsulated graphene devices. We additionally have identified a shift in the position of the 2D and G peaks with relative angle (fig. S2).

Mechanical resistance while pushing the BN imparted a torque to the AFM tip, causing it to cant away from a vertical position and produce a voltage difference in the AFM's photodetector. This effect can be used to identify variations in frictional forces (Fig. 2A). When sliding the top BN layer with the AFM tip in a translational motion far from alignment, we identified three regimes (Fig. 2B): (i) sliding friction between the tip and the substrate before the tip encountered the BN, (ii) static friction when the tip encountered the BN but resisted translation, and (iii) dynamic friction once the BN began to move.

Figure 2C shows a plot of the change in the tip deflection under continuous rotation (dynamic

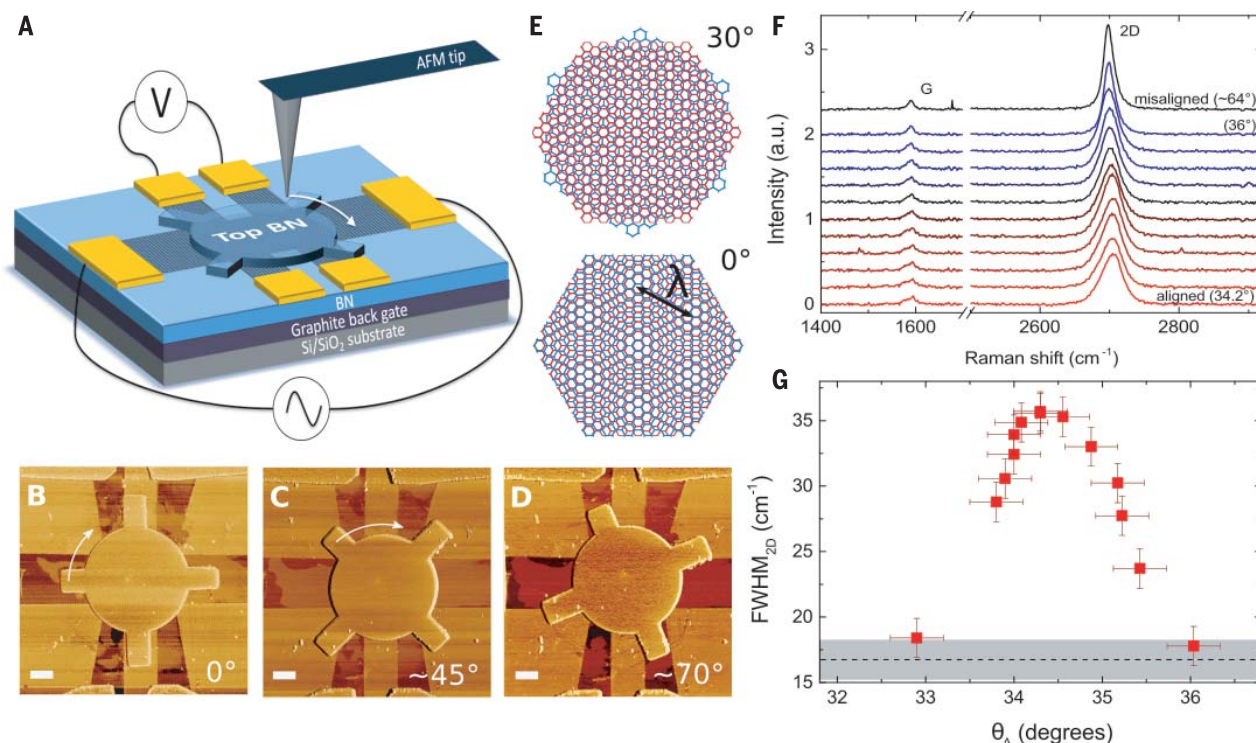
<sup>1</sup>Department of Physics, Columbia University, New York, NY, USA. <sup>2</sup>Department of Mechanical Engineering, Columbia University, New York, NY, USA. <sup>3</sup>Department of Electrical Engineering, Columbia University, New York, NY, USA.

<sup>4</sup>National Institute for Materials Science, 1-1 Namiki, Tsukuba, Japan.

\*These authors contributed equally to this work. †Present address: Centre de Nanosciences et de Nanotechnologies (C2N), CNRS, Université Paris Sud, Université Paris-Saclay, 91120 Palaiseau, France.

†Corresponding author. Email: rebeca.ribeiro@u-psud.fr (R.R.-P.); cd2478@columbia.edu (C.R.D.)



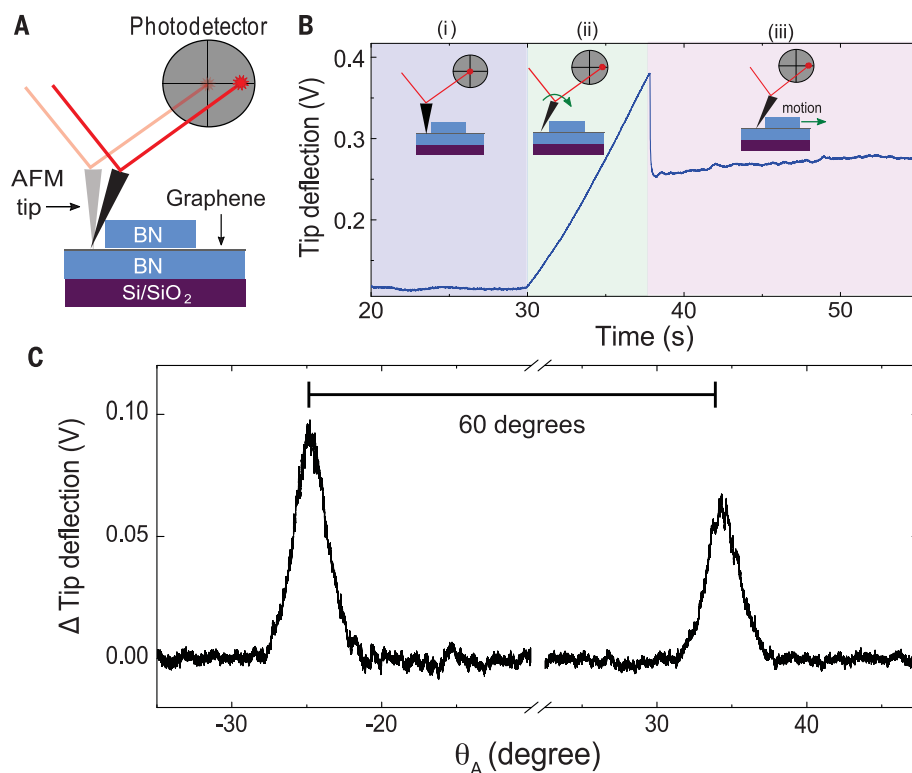


**Fig. 1. Rotatable heterostructures.** (A) Schematic cartoon of the device structure and the experimental technique. BN, hexagonal boron nitride. (B to D) AFM image of a fabricated device showing three different orientations of the top BN. The angle identified in each panel is the absolute angle referenced to the AFM coordinate system ( $\theta_A$ ). The images were acquired by the same AFM used to rotate the BN layer. Scale bars, 1  $\mu\text{m}$ . (E) Schematic illustration of the moiré superlattice arising between graphene (red) and BN (blue) at zero angle. The moiré wavelength,  $\lambda$ , is shown. (F) Raman spectrum of the device shown in

(B), (C), and (D) for  $\theta_A$  between  $34.2^\circ \pm 0.2^\circ$  and  $36.0^\circ \pm 0.2^\circ$ . The black curve shows an additional measurement acquired at  $\sim 64^\circ$  a.u., arbitrary units. (G) FWHM of the 2D peak as a function of the absolute angle. All Raman measurements were taken with the gate bias held at  $V_g = 0$  V. The peak FWHM position identifies zero-angle crystallographic alignment. Error bars represent the precision with which  $\theta$  can be determined from the AFM topographic images. The dashed line represents the FWHM measured for all angles larger than  $\sim 2^\circ$  away from perfect alignment, with the shaded region representing the associated uncertainty.

**Fig. 2. Mechanical properties.** (A) Schematic description of friction measurements.

When the AFM tip encountered the BN structure, it canted, causing a repositioning of the reflected laser spot in the four-quadrant photodetector. The resulting voltage difference was proportional to the tip cant angle (referred to here as the tip deflection) and served as a measure of the torque force acting at the end of the tip. (B) Tip deflection versus time in a translational push of the upper BN structure. Different regimes of the measurement were identified: (i) As the tip dragged along the surface, tip-substrate friction resulted in a steady-state tip deflection. (ii) When the tip encountered the BN structure, it initially resisted translation, and the tip deflection increased. We refer to this as the static friction regime. (iii) Once the BN was in motion, the tip deflection relaxed slightly, providing a measure of the dynamic friction at the BN-graphene interface. (C) Tip deflection versus absolute angle, measured during a continuous rotation of the BN. Two peaks were observed, spaced  $60^\circ$  apart.



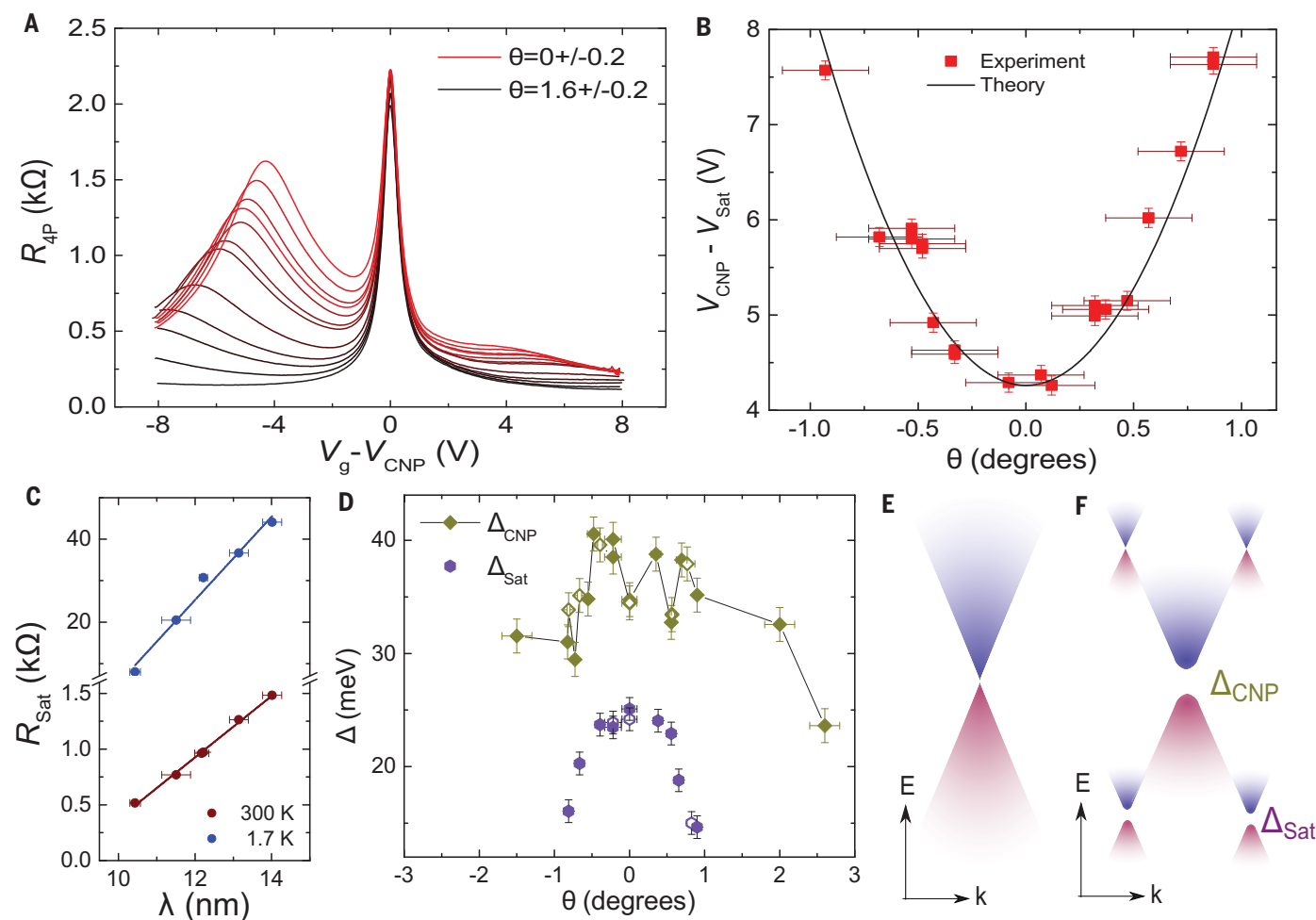
friction), in which we subtracted the background caused by piezoelectric drift and residual friction. Two prominent peaks are separated by  $60^\circ$ . This result closely resembles a previous measurement of friction between two graphitic structures in which a transition from superlubricity (where the structures are in an incommensurate position and the atomic shear forces are negligible) to a dissipative state was observed at commensurate angles of the threefold symmetric hexagonal lattices (30–32). However, because of the lattice mismatch between graphene and BN, there is not true lattice commensurability at any angle, and the increase in the friction should have a different origin. Recent numerical simulations suggest that contributions of the moiré superlattice to the frictional force cannot be neglected and are expected to be maximal for aligned layers (33), which could explain our experimental result. Our device structure allows us to study me-

chanical properties, such as frictional force, in atomically flat materials without rugosity contributions. These results also highlight the possibility of using the friction response as an in situ method to monitor and control layer alignment in heterostructures.

Our device design enabled us to measure electron transport in the active layer (in this case, graphene) while changing the relative orientation of the overlayer. Figure 3A shows a plot of the four-terminal resistance of the graphene layer as a function of back gate voltage  $V_g$  for different values of  $\theta$  at room temperature. Near  $\theta = 0$ , additional satellite resistance peaks appear symmetrically in density around the CNP. This result is consistent with the emergence of satellite Dirac points induced by scattering from the moiré superlattice potential (18, 20–22).

As  $\theta$  increased from zero, the satellite peaks diminished in intensity and moved further from

the CNP to higher gate values. To analyze this behavior more quantitatively, Fig. 3B plots the satellite peak position in  $V_g$  versus  $\theta$  determined from the AFM imaging. The measured position shows excellent agreement with the values of carrier density  $n$  at which the full filling of the miniband occurs,  $n = 8/\sqrt{3}\lambda^2$ , where the carrier density and gate voltage are related by  $n = C_g(V_g - V_{\text{CNP}})$ ,  $C_g$  is the capacitive coupling of the electrostatic gate, and  $\lambda$  is given by Eq. 1 (18, 20–22). The error bars in Fig. 3B,  $\sim \pm 0.2^\circ$ , reflect the precision with which  $\theta$  can be determined from the AFM topographic images. Determining  $\theta$  from the  $V_g$  position of the satellite peak at low temperature provides a more accurate measurement with uncertainty less than  $\pm 0.1^\circ$ . However, this method of determining the angle is limited to only a few degrees where the satellite peak remains within an accessible density range.

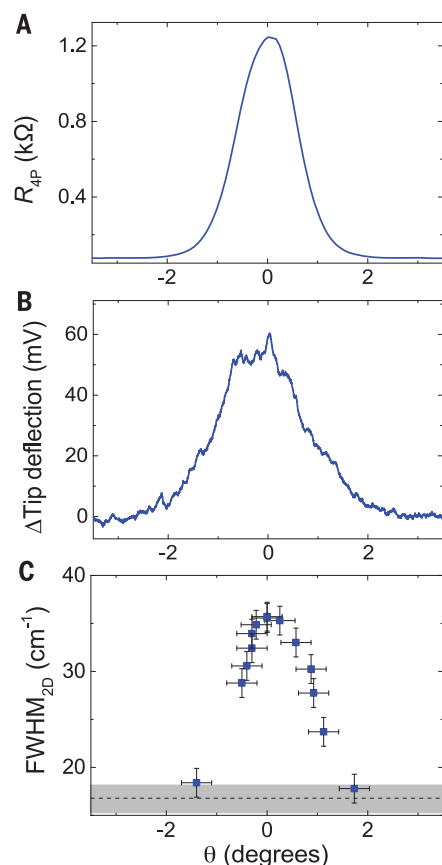


**Fig. 3. Electronic transport properties.** (A) Four-terminal resistance ( $R_{4P}$ ) as a function of  $V_g$  for different alignments of the graphene/BN structure, acquired at room temperature. (B) Position of the satellite peak in gate voltage as a function of the relative angle. The  $\pm 0.2^\circ$  error in the angle reports the precision achieved with the AFM imaging in tapping mode. The vertical error bars represent the maximum uncertainty with which we can determine the position of the satellite peak owing to the broadening of the peak. (C) Linear dependence of the maximum value of

the four-terminal resistance at the satellite peak at 300 K (red) and 1.7 K (blue) as a function of the moiré length. (D) Energy gap, measured by thermal activation, for the satellite peak (circles) and the charge neutrality point (diamonds) as a function of the relative angle. Open symbols represent a repeated measurement at a given angle after moving through other angles and thermally cycling. (E) Schematic band structure for native graphene and (F) for a graphene-BN heterostructure with a small twist angle.  $E$ , energy;  $k$ , momentum.

Plotting the resistance of the satellite peak versus  $\lambda$  for a single device revealed an apparently linear variation in the magnitude of the satellite peak resistance (Fig. 3C). The resistance of the CNP did not change linearly with  $\theta$  in this range (fig. S3). The origin of this linear dependence is not known. This striking observation highlights how a physical phenomenon previously obscured (18, 20–22) by sample-to-sample variations became clear by varying rotation in a single sample.

We measured the energy gap versus angle at both the central and satellite Dirac points by



**Fig. 4. Compilation of observations with the angle control technique.** (A) Four-terminal resistance as a function of the relative angle measured at a carrier density of  $-1.9 \times 10^{12} \text{ cm}^{-2}$ . (B) Tip deflection in friction measured simultaneously with the electronic transport. (C) FWHM of the 2D peak of the Raman spectrum as a function of the relative angle. All measurements were performed in the same device.

thermal activation. The gap magnitudes (Fig. 3D) were in good agreement with electronic transport measurements in encapsulated (34) and nonencapsulated devices (19, 20), optical measurements made in epitaxial BN/graphene structures (35), and theoretical calculations (36). The energy gap of the satellite peak decreased smoothly away from  $\theta = 0$ . In contrast, the energy gap at the CNP displayed a more complex behavior near  $\theta = 0$  and only decreased appreciably for  $|\theta| > 2^\circ$ .

The origin of the extra structure observed for the CNP remains to be understood. However, this result does not appear to be simple experimental noise because several gap values were confirmed to be reproducible when measured nonconsecutively (i.e., after thermally cycling and rotating through different angles and back). The difference in behavior of the two energy gaps, which would be impossible to observe in a study that used multiple samples at fixed angles (34), reflects their different physical origins and highlights the importance of these results in fully understanding the band structure modifications resulting from variations in angular alignment (Fig. 3, E and F). The persistence of an energy gap at the CNP in encapsulated devices for angles beyond that at which a commensurate-incommensurate transition was previously identified (19) suggests that this energy gap is not related solely to the presence of a commensurate state.

Figure 4 shows a direct comparison of the optical, mechanical, and electronic response versus angle measured in the same device. The four-probe resistance (Fig. 4A), the maximum tip deflection signal (friction; Fig. 4B), and the maximum FWHM of the Raman 2D peak (Fig. 4C) coincide exactly, confirming the relation between these properties. The four-probe resistance (Fig. 4A) and friction response (Fig. 4B) were acquired simultaneously at a fixed carrier density of  $1.9 \times 10^{12} \text{ cm}^{-2}$  (corresponding to the dashed line in fig. S4) while continuously rotating the BN layer. At this relatively high carrier density, the bulk resistance was modulated by more than an order of magnitude over  $<2^\circ$  of rotation (at cryogenic temperatures, this modulation increases to more than two orders of magnitude; supplementary materials).

Our demonstration of rotatable heterostructures with dynamically tunable device characteristics presents an opportunity in device engineering. Although we investigated BN-encapsulated graphene as a model tunable system, this technique is readily extendable to generic heterostructures fabricated from 2D materials. In addition to band structure tunability, emergent phases such as superconductivity and magnetism may be controllably varied with rotation.

Downloaded from <http://science.sciencemag.org/> on August 16, 2018

## REFERENCES AND NOTES

- S. Carr et al., *Phys. Rev. B* **95**, 075420 (2017).
- L. Britnell et al., *Nat. Commun.* **4**, 1794 (2013).
- A. Mishchenko et al., *Nat. Nanotechnol.* **9**, 808–813 (2014).
- M. T. Greenaway et al., *Nat. Phys.* **11**, 1057–1062 (2015).
- B. Fallahzad et al., *Nano Lett.* **15**, 428–433 (2015).
- T. Chari, R. Ribeiro-Palau, C. R. Dean, K. Shepard, *Nano Lett.* **16**, 4477–4482 (2016).
- E. Koren et al., *Nat. Nanotechnol.* **11**, 752–757 (2016).
- J. R. Wallbank et al., *Science* **353**, 575–579 (2016).
- L. A. Gonzalez-Arraga, J. L. Lado, F. Guinea, P. San-Jose, *Phys. Rev. Lett.* **119**, 107201 (2017).
- Y. Cao et al., *Phys. Rev. Lett.* **117**, 116804 (2016).
- J. D. Sanchez-Yamagishi et al., *Nat. Nanotechnol.* **12**, 118–122 (2017).
- L. Ju et al., *Nature* **520**, 650–655 (2015).
- C.-J. Kim et al., *Nat. Nanotechnol.* **11**, 520–524 (2016).
- Y. Cao et al., *Nature* **556**, 43–50 (2018).
- H. Yu, Y. Wang, Q. Tong, X. Xu, W. Yao, *Phys. Rev. Lett.* **115**, 187002 (2015).
- P. Rivera et al., *Science* **351**, 688–691 (2016).
- P. Rivera et al., *Nat. Commun.* **6**, 6242 (2015).
- M. Yankowitz et al., *Nat. Phys.* **8**, 382–386 (2012).
- C. R. Woods et al., *Nat. Phys.* **10**, 451–456 (2014).
- B. Hunt et al., *Science* **340**, 1427–1430 (2013).
- L. A. Ponomarenko et al., *Nature* **497**, 594–597 (2013).
- C. R. Dean et al., *Nature* **497**, 598–602 (2013).
- K. Kim et al., *Nano Lett.* **16**, 1989–1995 (2016).
- D. Wang et al., *Phys. Rev. Lett.* **116**, 126101 (2016).
- C. R. Woods et al., *Nat. Commun.* **7**, 10800 (2016).
- L. Wang et al., *Science* **342**, 614–617 (2013).
- C. R. Dean et al., *Nat. Nanotechnol.* **5**, 722–726 (2010).
- A. Eckmann et al., *Nano Lett.* **13**, 5242–5246 (2013).
- C. Neumann et al., *Nat. Commun.* **6**, 8429 (2015).
- M. Dienwiebel et al., *Phys. Rev. Lett.* **92**, 126101 (2004).
- Z. Liu et al., *Phys. Rev. Lett.* **108**, 205503 (2012).
- A. E. Filippov, M. Dienwiebel, J. W. M. Frenken, J. Klafter, M. Urbakh, *Phys. Rev. Lett.* **100**, 046102 (2008).
- E. Koren, U. Duerig, *Phys. Rev. B* **94**, 045401 (2016).
- L. Wang et al., *Science* **350**, 1231–1234 (2015).
- Z.-G. Chen et al., *Phys. Commun.* **5**, 4461 (2014).
- J. C. W. Song, A. V. Shytov, L. S. Levitov, *Phys. Rev. Lett.* **111**, 266801 (2013).

## ACKNOWLEDGMENTS

We acknowledge discussions with M. Yankowitz and A. Botello-Mendez, and we acknowledge J. Huerta, N. Finney, T. Chari, S. Chen, and M. Gustafsson for technical support. **Funding:** This research was supported by the NSF MRSEC program through Columbia University's Center for Precision Assembly of Superstratic and Superatomic Solids (DMR-1420634). C.R.D. acknowledges partial support by the National Science Foundation (DMR-1462383). **Author contributions:** R.R.-P. and C.R.D. designed the experiment. R.R.-P. and C.Z. fabricated the samples, performed the experiments, analyzed the data, and wrote the paper. T.T. and K.W. grew the crystals of hexagonal boron nitride. J.H. and C.R.D. advised on experiments, data analysis, and writing the paper. **Competing interests:** None declared. **Data and materials availability:** All data needed to evaluate the conclusions are present in the paper or the supplementary materials.

## SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/361/6403/690/suppl/DC1](http://www.sciencemag.org/content/361/6403/690/suppl/DC1)  
Materials and Methods  
Figs. S1 to S9  
References (37, 38)

6 April 2018; accepted 11 June 2018  
10.1126/science.aat6981



## PLANT SCIENCE

# Plant iron acquisition strategy exploited by an insect herbivore

L. Hu<sup>1</sup>, P. Mateo<sup>1,2</sup>, M. Ye<sup>1</sup>, X. Zhang<sup>1</sup>, J. D. Berset<sup>1</sup>, V. Handrick<sup>3\*</sup>, D. Radisch<sup>3</sup>, V. Grabe<sup>3</sup>, T. G. Köllner<sup>3</sup>, J. Gershenzon<sup>3</sup>, C. A. M. Robert<sup>1†</sup>, M. Erb<sup>1†</sup>

Insect herbivores depend on their host plants to acquire macro- and micronutrients. Here we asked how a specialist herbivore and damaging maize pest, the western corn rootworm, finds and accesses plant-derived micronutrients. We show that the root-feeding larvae use complexes between iron and benzoxazinoid secondary metabolites to identify maize as a host, to forage within the maize root system, and to increase their growth. Maize plants use these same benzoxazinoids for protection against generalist herbivores and, as shown here, for iron uptake. We identify an iron transporter that allows the corn rootworm to benefit from complexes between iron and benzoxazinoids. Thus, foraging for an essential plant-derived complex between a micronutrient and a secondary metabolite shapes the interaction between maize and a specialist herbivore.

Iron (Fe) can be a limiting micronutrient for plants and herbivores (1). Plants increase Fe availability by secreting reducing agents and Fe chelators, so-called phytosiderophores, into the rhizosphere (2). Benzoxazinoid secondary metabolites, which are produced and exuded by grasses such as maize, may act as phytosiderophores in addition to protecting plants against herbivores (3, 4). The western corn rootworm (WCR), a worldwide maize pest, can tolerate and sequester benzoxazinoids (5). In addition, root-feeding WCR larvae prefer benzoxazinoid-producing maize plants and require benzoxazinoids to identify crown roots as the preferred feeding site (6). WCR crown root damage reduces plant growth (6).

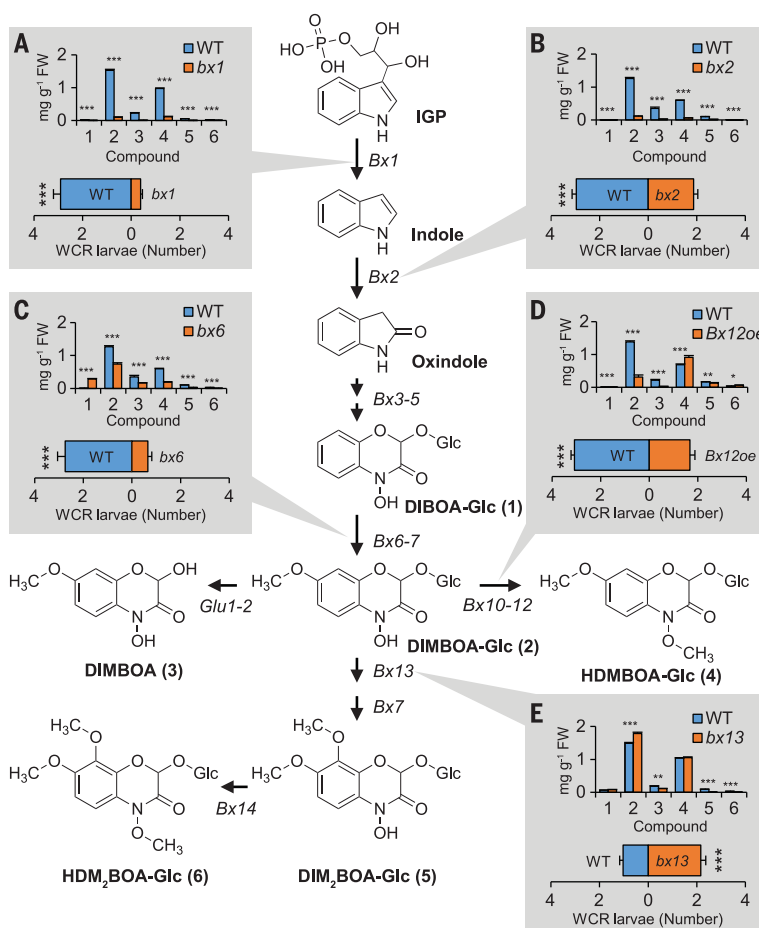
Here, we investigate the potential role of benzoxazinoids as phytosiderophores and WCR foraging cues. To identify the chemical motif used by WCR, we evaluated WCR behavior on different maize benzoxazinoid mutants (Fig. 1 and fig. S1). WCR larvae preferred wild-type (WT) rather than benzoxazinoid-deficient *bx1* and *bx2* mutant plants. The larvae also preferred WT to *bx6* mutants, which overaccumulate DIBOA-Glc at the

expense of 7-*O*-methylated and 8-*O*-methylated benzoxazinoids, suggesting that DIBOA-Glc is not a preferred benzoxazinoid. WCR larvae preferred *bx13* mutant plants deficient in 8-*O*-methylated benzoxazinoids, suggesting that 8-*O*-methylated benzoxazinoids are not preferred. To test the influence of *N*-*O*-methylation, we over-

expressed the DIMBOA-Glc *O*-methyltransferase *ZmBx12* (*Bx12oe*) (fig. S1), resulting in plants with excess HDMBOA-Glc and reduced DIMBOA and DIMBOA-Glc (Fig. 1). WCR preferred WT over *ZmBx12*-overexpressing plants, suggesting that *N*-*O*-methylated benzoxazinoids are not preferred (Fig. 1). Thus, 7-*O*-methylated, *N*-hydroxylated benzoxazinoids such as DIMBOA and DIMBOA-Glc are associated with WCR feeding preference.

We next investigated the role of benzoxazinoids for within-plant feeding preferences. WCR no longer distinguished between crown and primary roots of *bx1*, *bx2*, *bx6*, and *Bx12oe* plants, while preference was intact in *bx13* plants (figs. S2 and S3). The correlation between plant preference and tissue preference suggests that the same benzoxazinoids may mediate plant preference and within-plant foraging. Closer inspection of crown and primary root benzoxazinoid profiles revealed a positive correlation between DIMBOA accumulation in crown roots relative to primary roots and WCR preference for crown roots relative to primary roots across the different maize mutants ( $R^2 = 0.56$ ;  $P = 0.03$ , fig. S2).

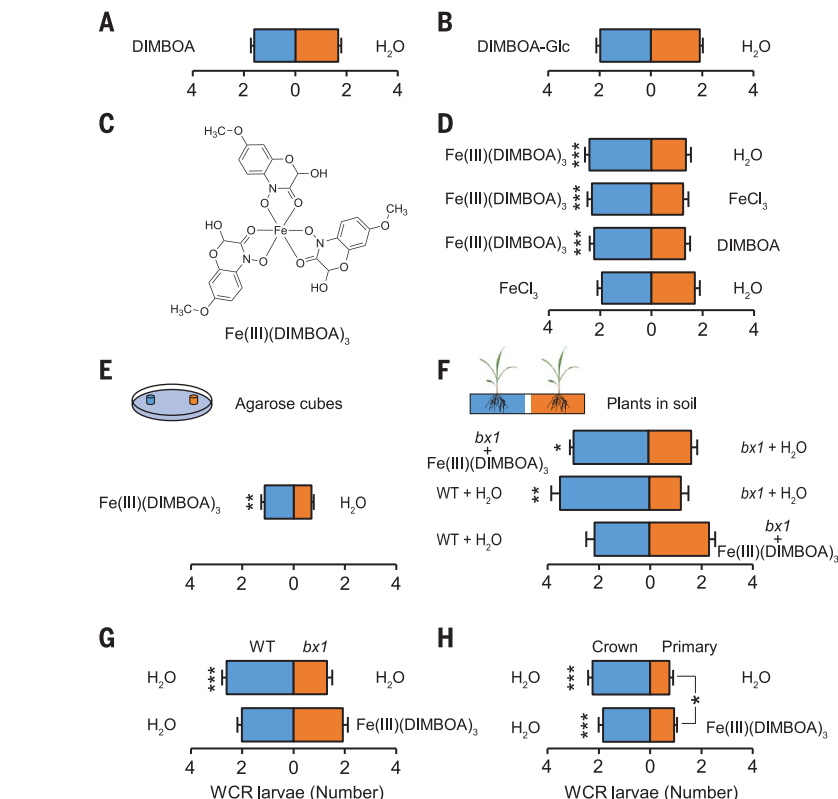
Contrary to what was expected from the experiments with the benzoxazinoid mutant plants, purified DIMBOA and DIMBOA-Glc did not elicit WCR feeding preference (Fig. 2). The DIMBOA breakdown product MBOA, a potential volatile WCR attractant (7), was equally inactive (fig. S4).



Because *N*-hydroxylated benzoxazinoids such as DIMBOA and DIMBOA-Glc, but not *N*-O-methylated benzoxazinoids, can form stable complexes with Fe (3), we hypothesized that these complexes may serve as WCR foraging cues. Within the pH range of the maize rhizosphere (4.5 to 7.5) and given that DIMBOA is the main benzoxazinoid around maize roots (8), Fe complexes with two [referred to as Fe(III)(DIMBOA)<sub>2</sub>] and three DIMBOA ligands [referred to as Fe(III)(DIMBOA)<sub>3</sub>] are likely to accumulate at the root surface. Spectrophotometric measurements of synthetically prepared complexes (fig. S5A) confirmed the pH-dependent formation of Fe(III)(DIMBOA)<sub>2</sub> and Fe(III)(DIMBOA)<sub>3</sub> (collectively referred to as Fe-DIMBOA, fig. S5B). DIMBOA reacted preferentially with Fe, followed by aluminum (Al) and molybdenum (Mo) (fig. S6). Formation of Fe-DIMBOA was observed on WT crown roots, but not on *bx1* roots (fig. S7). At estimated physiological doses of  $1.26 \times 10^{-10}$  mol cm<sup>-2</sup> (fig. S8), Fe(III)(DIMBOA)<sub>2</sub> and Fe(III)(DIMBOA)<sub>3</sub>, but not Al-DIMBOA, Fe-DIMBOA-Glc, and Fe-EDTA, elicited WCR feeding preference (Fig. 2D and fig. S9). Fe(III)(DIMBOA)<sub>3</sub> increased WCR feeding across a range of physiologically relevant concentrations from  $1.26 \times 10^{-14}$  to  $1.26 \times 10^{-10}$  mol cm<sup>-2</sup> (fig. S10). Preference for Fe(III)(DIMBOA)<sub>3</sub> was independent of larval stage, previous benzoxazinoid exposure, and the presence of maize roots (Fig. 2E and fig. S11). Fe(III)(DIMBOA)<sub>3</sub> rescued recruitment of WCR larvae to *bx1* mutants in soil (Fig. 2F) and to individual *bx1* roots (Fig. 2G). Primary roots of WT plants became more attractive upon Fe(III)(DIMBOA)<sub>3</sub> application and were as attractive as crown roots 6 hours after application (Fig. 2H and fig. S12). Host plant acceptance on *bx1* mutant roots was fully complemented by Fe(III)(DIMBOA)<sub>3</sub> (fig. S13). Fe(III)(DIMBOA)<sub>3</sub> application also increased WCR feeding on rice and barley, two non-host plant species (fig. S13). Thus, in addition to primary metabolites (9, 10), Fe-DIMBOA mediates host recognition, acceptance, and within-plant foraging of WCR.

To investigate the impact of Fe-DIMBOA on WCR performance, we grew WT and *bx1* mutant plants in nutrient solutions. In the presence of free Fe that can form complexes with DIMBOA, WCR larvae grew better on WT than *bx1* plants (Fig. 3A). No difference was observed in the presence of Fe-EDTA or in the absence of Fe. Exogenous application of DIMBOA rescued larval growth on *bx1* mutants (Fig. 3B). Fe concentrations in WCR mirrored larval performance (Fig. 3C). Thus, the interaction between free Fe and DIMBOA increases WCR Fe supply and performance.

Fe supply is critical for plant performance. Leaf chlorosis, a typical sign of iron deficiency, was observed in *bx1*, *bx2*, and *bx6* mutant seed-



**Fig. 2. Complexes between iron (Fe) and the benzoxazinoid DIMBOA mediate host recognition, acceptance, and within-plant foraging of WCR larvae.** (A and B) Influence of pure DIMBOA and DIMBOA-Glc on WCR feeding preference on benzoxazinoid-deficient mutant roots. (C) Chemical structure of the Fe complex Fe(III)(DIMBOA)<sub>3</sub>. (D and E) Effect of synthetic Fe(III)(DIMBOA)<sub>3</sub> on WCR feeding preference for roots (D) and agarose cubes (E). (F) WCR feeding preference for soil-grown WT and *bx1* mutant plants supplied with Fe(III)(DIMBOA)<sub>3</sub>. (G and H) Influence of Fe(III)(DIMBOA)<sub>3</sub> complementation on within- and between-plant WCR feeding preferences. [+SE, *n* = 20, except (F), *n* = 10 to 11 choice situations with five larvae each]. Full time courses are shown in figs. S4, S9, and S12. Asterisks indicate significant differences (\**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001).

lings growing in nutrient solution and potting soil supplemented with Fe salts (Fig. 3E and fig. S14). No chlorosis was observed in *bx* mutants supplied with Fe-EDTA (fig. S14). DIMBOA supplementation rescued *bx1* mutants supplied with free Fe (fig. S14). Fe concentrations in the maize xylem mirrored chlorophyll patterns (Fig. 3, F and G). No differences in plant biomass were observed between genotypes (fig. S14F). Thus, the interaction between free Fe and DIMBOA increases maize Fe supply. As maize and WCR Fe contents are correlated across treatments and genotypes, the positive effect of Fe-DIMBOA on WCR may be due to direct or plant-mediated effects.

To better understand the connection between WCR performance and Fe availability, we manipulated the capacity of WCR to acquire Fe. We identified a WCR homolog of the human divalent metal transporter-1 (DMT1) (11), here named *DvIRT1* (fig. S15). In *Drosophila melanogaster*, the DMT1 homolog *Mtv* is required for Fe homeostasis and feeding decisions (12). *DvIRT1* rescued the growth of an Fe-transport-deficient yeast strain in the presence of free or complexed Fe, including Fe-DIMBOA (Fig. 4A). Silencing of

*DvIRT1* in WCR (Fig. 4B) resulted in WCR Fe deficiency (Fig. 4C). *DvIRT1* silencing did not change WCR feeding preferences (Fig. 4D). However, *DvIRT1* was required for the benzoxazinoid-dependent increase in WCR Fe supply and performance (Fig. 4E). Thus, *DvIRT1* enables WCR to acquire Fe in various forms, including Fe-DIMBOA.

WCR may derive multiple benefits from Fe-DIMBOA. First, as Fe-DIMBOA is only produced by a few other plant species (13), it is a reliable host-recognition cue. Second, Fe-DIMBOA levels are highest for crown roots, which are a better food source for WCR than primary roots (5). Third, Fe-DIMBOA is accepted as a substrate by *DvIRT1* and directly improves Fe homeostasis and WCR performance. Fourth, WCR larvae sequester the DIMBOA breakdown product MBOA for self-defense against entomopathogenic nematodes (6). Fe-DIMBOA may therefore provide Fe as well as DIMBOA as an immune precursor.

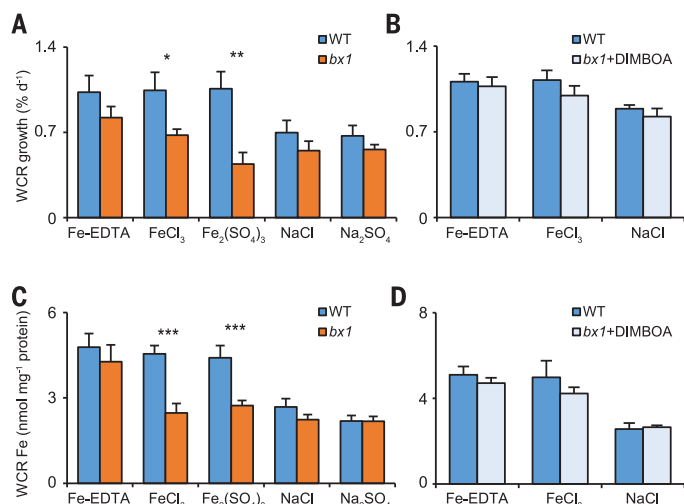
Grasses can use L-methionine-derived mugenic acids to chelate Fe (2). Here, we show that benzoxazinoids also contribute to the Fe supply of young maize plants. In addition, benzoxazinoids

<sup>1</sup>Institute of Plant Sciences, University of Bern, Switzerland.

<sup>2</sup>Neuchâtel Platform of Analytical Chemistry, University of Neuchâtel, Switzerland. <sup>3</sup>Max Planck Institute for Chemical Ecology, Jena, Germany.

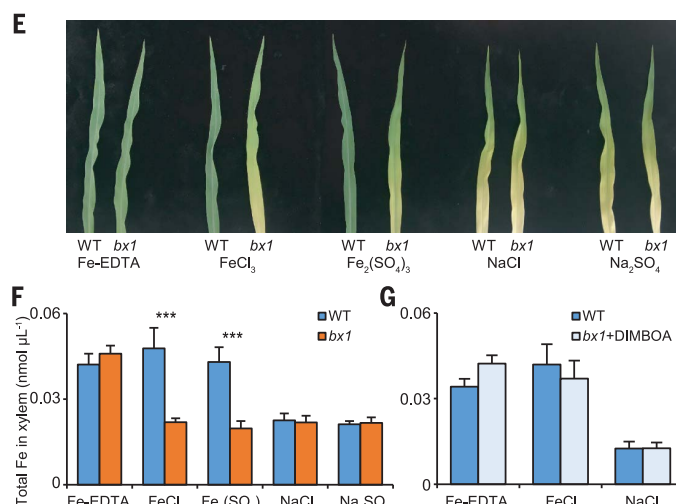
\*Present address: John Innes Centre, Norwich, UK.

†Corresponding author. Email: christelle.robert@ips.unibe.ch (C.A.M.R.); matthias.erb@ips.unibe.ch (M.E.)

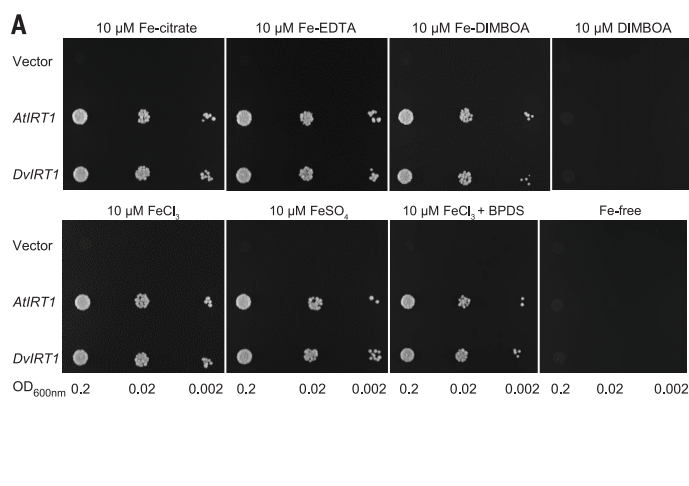


**Fig. 3. Interactions between free Fe and the benzoxazinoid DIMBOA determine growth and Fe homeostasis of maize and WCR larvae.**

(A and B) Growth of WCR larvae feeding on WT or *bx1* roots supplied with different sources of Fe (+SE,  $n = 9$  to 15 biological replicates) and pure DIMBOA (+SE,  $n = 15$  to 20 biological replicates). (C and D) Corresponding Fe contents of WCR (+SE,  $n = 5$  biological replicates,

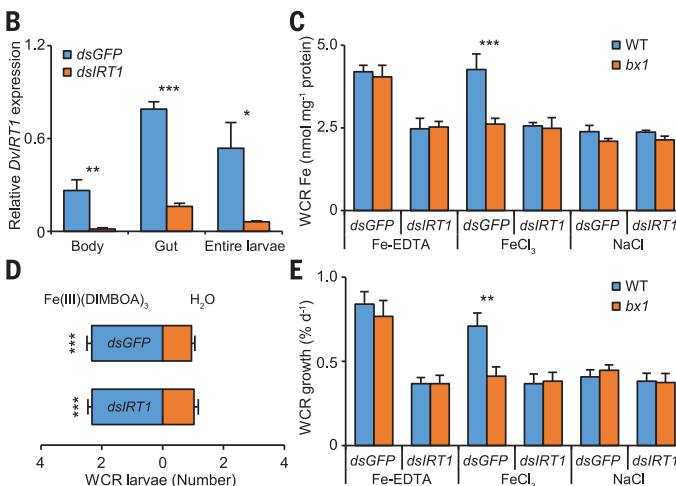


with 4 to 5 larvae pooled per replicate). (E) Representative photographs of leaves of WT and *bx1* maize plants grown in nutrient solutions with different sources of Fe. (F and G) Average Fe content in the maize xylem sap (+SE,  $n = 5$  biological replicates, with sap of four plants pooled per replicate). Asterisks indicate significant differences ( $*P < 0.05$ ;  $**P < 0.01$ ;  $***P < 0.001$ ).



**Fig. 4. The Fe transporter DvIRT1 is required for Fe-DIMBOA-dependent WCR performance.**

(A) Functional complementation of a yeast Fe-uptake mutant. Plasmids expressing *AtIRT1* (positive control, *DvIRT1*) or the empty vector pFL61 (negative control) were individually introduced into the Fe uptake-defective yeast DEY1453 strain, and the growth media were supplemented with different Fe sources. (B) Average expression levels of *DvIRT1* in WCR larvae after feeding with double-stranded RNA of *GFP*



(*dsGFP*, negative control) or *DvIRT1* (*dsIRT1*, +SE,  $n = 3$  biological replicates). (D) Feeding preference of *dsGFP*- and *dsIRT1*-fed WCR larvae (+SE,  $n = 19$  to 20 choice situations with five larvae each). (C and E) Fe content (+SE,  $n = 4$  biological replicates, with 3 to 5 larvae pooled per replicate) and growth (+SE,  $n = 20$  biological replicates) of *dsGFP*- or *dsIRT1*-exposed larvae feeding on WT or *bx1* roots with different Fe source treatments. Full time courses are shown in fig. S16. Asterisks indicate significant differences ( $*P < 0.05$ ;  $**P < 0.01$ ;  $***P < 0.001$ ).

also act as resistance factors against different pests and diseases and shape the root microbiome (8, 14) but, as shown here, also increase the performance of a specialist herbivore. The diverse costs and benefits of the benzoxazinoid pathway for maize represent an optimization problem for plant breeding that may have contributed to the persistence of WCR as a damaging maize pest.

Essential trace metals such as Fe influence herbivore performance and herbivore community composition (15–17). As trace metals are

often present as complexes in plants (2, 18), the ability of herbivores to detect and respond to these complexes may shape plant-herbivore interactions in agricultural and natural ecosystems.

#### REFERENCES AND NOTES

- J. F. Briat, C. Dubos, F. Gaymard, *Trends Plant Sci.* **20**, 33–40 (2015).
- T. Kobayashi, N. K. Nishizawa, *Annu. Rev. Plant Biol.* **63**, 131–152 (2012).
- L. Bigler, A. Baumbach, C. Werner, M. Hesse, *Helv. Chim. Acta* **79**, 1701–1709 (1996).

- S. Zhou, A. Richter, G. Jander, *Plant Cell Physiol.* 10.1093/pcp/pcy064 (2018).
- C. A. M. Robert et al., *Ecol. Lett.* **15**, 55–64 (2012).
- C. A. M. Robert et al., *eLife* **6**, e29307 (2017).
- L. B. Bjostad, B. E. Hibbard, *J. Chem. Ecol.* **18**, 931–944 (1992).
- L. Hu et al., *Nat. Commun.* **9**, 2738 (2018).
- E. J. Bernklau, L. B. Bjostad, *J. Econ. Entomol.* **101**, 341–351 (2008).
- E. J. Bernklau, B. E. Hibbard, D. L. Dick, C. D. Rithner, L. B. Bjostad, *J. Econ. Entomol.* **108**, 539–548 (2015).
- H. Gunshin et al., *Nature* **388**, 482–488 (1997).
- V. Rodrigues, P. Y. Cheah, K. Ray, W. Chia, *EMBO J.* **14**, 3007–3020 (1995).
- M. Frey, K. Schullehner, R. Dick, A. Fiessmann, A. Gierl, *Phytochemistry* **70**, 1645–1651 (2009).



14. F. C. Wouters, B. Blanchette, J. Gershenzon, D. G. Vassão, *Phytochem. Rev.* **15**, 1127–1151 (2016).
15. J. L. Freeman, C. F. Quinn, M. A. Marcus, S. Fakra, E. A. H. Pilon-Smits, *Curr. Biol.* **16**, 2181–2192 (2006).
16. A. Joern, T. Provin, S. T. Behmer, *Ecology* **93**, 1002–1015 (2012).
17. A. Kazemi-Dinan, S. Thomaschky, R. J. Stein, U. Krämer, C. Müller, *New Phytol.* **202**, 628–639 (2014).
18. E. Andresen, E. Peiter, H. Küpper, *J. Exp. Bot.* **69**, 909–954 (2018).

#### ACKNOWLEDGMENTS

We thank W. French and C. Nielson (USDA Brookings) for supplying WCR eggs, D. Eide and J. Taggart (University of

Wisconsin–Madison) for sharing yeast strains and plasmids, and N. Ségaud and M. Albrecht (University of Bern) for help with Fe-DIMBOA characterization. **Funding:** The work of M.E. was supported by the Swiss National Science Foundation (grants 155781, 160786, 157884) and the University of Bern.

**Author contributions:** Study design: M.E., C.A.M.R. Experimental design and supervision: L.H., M.E., C.A.M.R., T.G.K., J.G. Data collection and analysis: L.H., C.A.M.R., P.M., M.Y., X.Z., J.D.B., V.H., D.R., V.G., T.G.K., M.E. Initial draft: L.H., M.E., C.A.M.R. Final version: All authors. **Competing interests:** The authors declare that they have no competing interests.

**Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the main text or the supplementary materials. The sequence of *DvIRT1* has been

deposited at NCBI (accession number MH715476). Additional data related to this paper may be requested from the authors.

#### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/361/6403/694/suppl/DC1](http://www.sciencemag.org/content/361/6403/694/suppl/DC1)

Materials and Methods

Figs. S1 to S19

Table S1

Data File S1

References (19–56)

23 February 2018; accepted 19 June 2018  
10.1126/science.aat4082

## CLIMATE

# Abrupt cloud clearing of marine stratocumulus in the subtropical southeast Atlantic

Sandra E. Yuter<sup>1\*</sup>, John D. Hader<sup>1,2</sup>, Matthew A. Miller<sup>1</sup>, David B. Mechem<sup>3</sup>

We document rapid and abrupt clearings of large portions of the subtropical marine low cloud deck that have implications for the global radiation balance and climate sensitivity. Over the southeast Atlantic, large areas of stratocumulus are quickly eroded, yielding partial or complete clearing along sharp transitions hundreds to thousands of kilometers in length that move westward at 8 to 12 meters per second and travel as far as 1000+ kilometers from the African coast. The westward-moving cloudiness reductions have an annual peak in occurrence in the period from April through June. The cloud erosion boundaries reduce cloud at  $\approx 10$ -kilometer scale in less than 15 minutes, move approximately perpendicular to the mean flow, and are often accompanied by small-scale wave features. Observations suggest that the cloud erosion is caused by atmospheric gravity waves.

The areal extent and temporal variability of subtropical marine low clouds strongly influence the global radiation balance (1). Large, persistent areas of subtropical marine stratocumulus clouds have been called Earth's "climate refrigerator" (2). These low marine clouds scatter solar radiation back to space and emit thermal radiation at a temperature close to the sea surface temperature (SST), yielding net radiative cooling of the climate system. The controls on areal coverage of marine stratocumulus have historically been cast either as a steady-state response to an imposed large-scale mean forcing (e.g., SST, subsidence, inversion strength, and radiative flux) (3, 4) or as being dominated by internal aerosol-cloud-precipitation processes (5–7). Most previous work on low marine cloudiness transitions, including several studies addressing pockets of open cells, has predominantly used frameworks that focus on responses to SST gradients and/or internal aerosol-cloud-precipitation processes and implicitly exclude external multiday synoptic variability (8, 9). Yet a recent spectral analysis of multiyear global satellite datasets (10) found that cloudiness variance at multiday (3 to 50 days) time scales exceeds the seasonal variance over the subtropical southeast Atlantic and northeast Pacific and is only slightly lower than the seasonal magnitude over the subtropical southeast Pacific. Abrupt changes in regional albedo from 0.9 (mostly low cloud cover) to 0.6 (ocean under mostly clear sky) will substantially increase the shortwave radiation absorbed by the ocean, which has pronounced implications for marine ecology and biogeochemistry (11).

Cloudiness transition boundaries hundreds to thousands of kilometers long in the southeast

Atlantic exhibit abrupt reductions of cloud akin to pulling away a sun shade (Fig. 1 and movies S1 to S4). The information below is derived from 377 cloud erosion boundaries, identified from an examination of 1911 days over the period from 8 May 2012 through 31 July 2017 by using moderate resolution imaging spectroradiometer (MODIS) corrected reflectance data and geosynchronous satellite visible and infrared (IR) data (see methods for details). The cloud-eroding events in the southeast Atlantic occur year-round, with a peak monthly frequency of occurrence of roughly 20 events per month in May (Fig. 2). The abrupt cloudiness transitions, up to 1000+ km in meridional length, move westward at 8 to 12 m s<sup>-1</sup> from 11°E to as far as 4°W longitude, as much as 1500 km from the African coastline (Fig. 3). For any single cloudiness boundary, the speed of motion is relatively constant as the boundary moves westward (Fig. 3). At different locations along the same boundary, cloud reduction can result in partial or complete clearing (Fig. 1). Cloudiness transitions typically become discernible along the coast of southwestern Africa, generally within a few hours of local midnight (Fig. 3 and movies S2 and S4). The removal of cloud at night indicates that shortwave radiation–cloud feedbacks are not required. The prevailing winds at cloud level in this region are typically southerly or southeasterly (12), and individual features in the low cloud field can be tracked moving northward relative to the westward-moving cloud erosion boundary (movie S1). The motion of the cloud erosion boundary is roughly perpendicular to the cloud-level winds and cannot be explained primarily by advection.

High-resolution imagery of areas along the cloudiness boundaries often show a sharp and abrupt transition from overcast to clear or from overcast to broken clouds over spatial scales of  $\leq 10$  km (Fig. 1). Along some cloudiness boundaries, individual closed cells within the mesoscale cellular structure of the marine stratocumulus

appear to be bisected (Fig. 1, F and G). Animated loops of visible satellite imagery reveal that cloud cover over  $\approx 10$ -km horizontal scales is removed or reduced in 15 min or less (movies S1 and S3). Near the cloud erosion boundaries, sets of wavelike, narrow (wavelength  $\leq 10$  km), elongated, banded features roughly parallel to the cloudiness boundaries frequently occur (Fig. 1, E and G).

Globally, satellite imagery reveals a rich array of multiday marine low-cloudiness variations, including sharp transitions between areas with nearly overcast and nearly clear conditions. Some of these cloudiness transition boundaries are associated with the advection of low-level air from poleward extratropical cyclones into the subtropics (13–15). Cloud coverage, type, and height in the midlatitudes are clearly modulated by synoptic-scale baroclinic weather systems (16–18). Some subtropical cloudiness reductions are associated with inversion strength anomalies originating from higher-latitude synoptic storms (10). Aircraft measurements off the coast of California show that near-shore clearings of the marine cloud field can be associated with synoptic-scale perturbations in the alignment and strength of the northeast Pacific ridge and concurrent mesoscale circulations along the California coast (19). These near-shore clearings expand during the day and contract toward the coast at night.

Propagating atmospheric gravity waves (buoyancy oscillations) can manifest as transitions from overcast to broken cloud conditions and from thicker to thinner clouds. A westward-moving diurnal atmospheric gravity wave originating over the heated, elevated terrain of the Andes modulates the liquid water path (LWP) in marine low clouds over the southeast Pacific (20–22). Intermittent atmospheric gravity wave trains originating from a disturbed subtropical jet can move equatorward into the southeast Pacific subtropical cloud deck, modulate the LWP, and reduce the cloud fraction during the day (23, 24).

The characteristics of the southeast Atlantic cloud clearings are inconsistent with advection and the specific gravity-wave mechanisms documented in the southeast Pacific. The westward motion of the southeast Atlantic cloudiness transitions in an environment of prevailing southerly and southeasterly cloud-level winds and a persistent, large-scale stable layer topped by an inversion that can serve as a waveguide (25) strongly suggest that the cloud erosion is caused primarily by an atmospheric gravity wave rather than advection. Additionally, the combination of the direction of motion, the phase speed, and the timing characteristics of these wavelike phenomena over the subtropical southeast Atlantic differs from that of documented atmospheric gravity waves over the subtropical southeast Pacific. The westward-moving diurnal solitary wave over the southeast Pacific is associated with heating of elevated terrain and crosses the coast at about 1700 local time (20, 21). This westward-propagating atmospheric gravity wave has a wavelength of  $\approx 400$  km, a phase

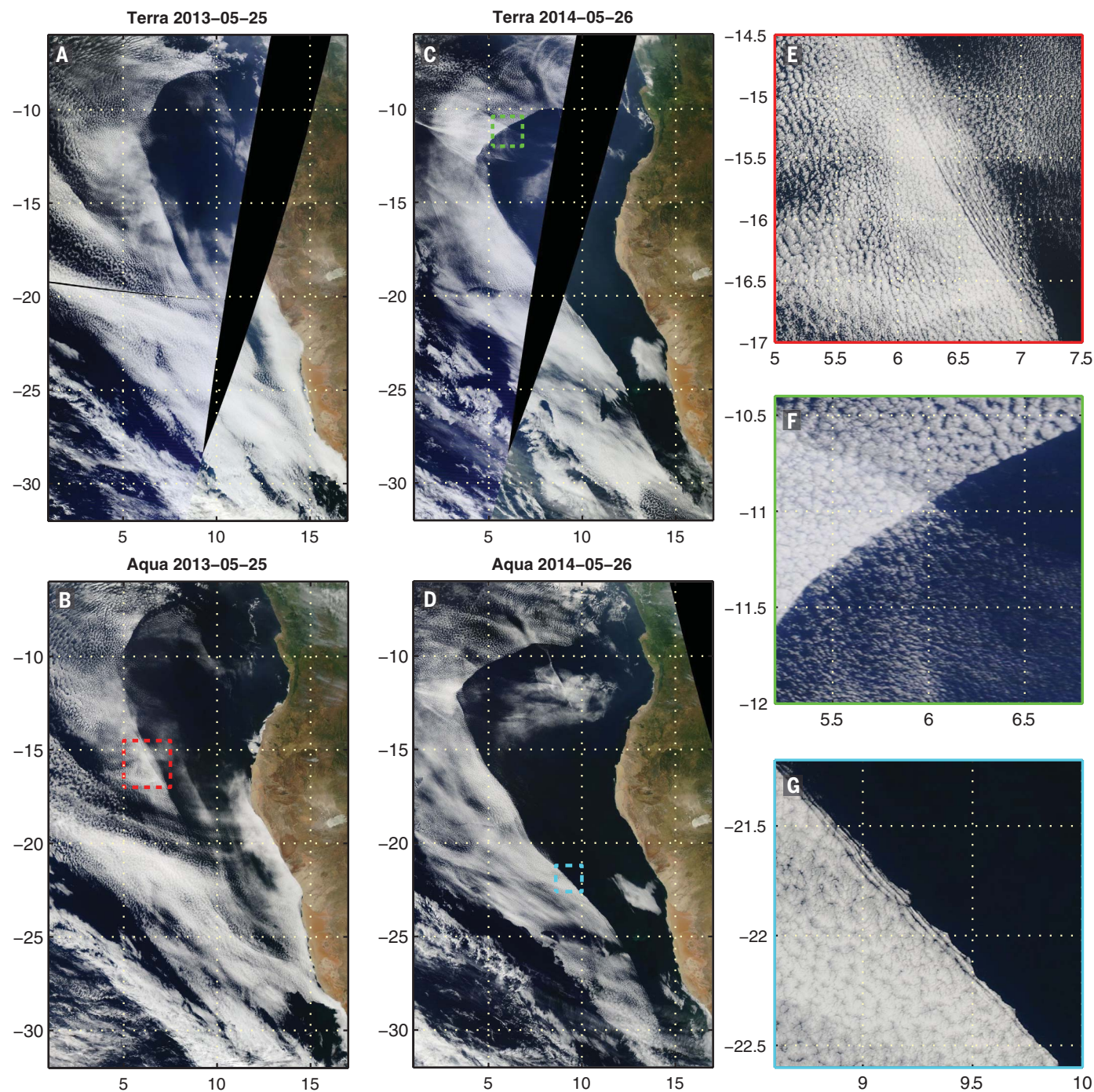
<sup>1</sup>Department of Marine, Earth, and Atmospheric Sciences, North Carolina State University, Raleigh, NC, USA. <sup>2</sup>ICF, Fairfax, VA, USA. <sup>3</sup>Department of Geography and Atmospheric Science, University of Kansas, Lawrence, KS, USA.

\*Corresponding author. Email: seyuter@ncsu.edu

speed of  $30 \text{ m s}^{-1}$ , and a wave depth of 5 km. The northeastward-moving atmospheric gravity wave trains that emerge from a disturbed subtropical jet and propagate toward the coast of Peru (23, 24) have wavelengths of 50 to 100 km, a vertical displacement of  $\approx 400 \text{ m}$ , and a phase speed of  $\approx 15 \text{ m s}^{-1}$ . The wavelength and vertical dis-

placement of the  $>1000\text{-km}$ -long southeast Atlantic cloud-eroding waves are currently not known, though the phase speed is in the range of 8 to  $12 \text{ m s}^{-1}$  and evidence of waves with wavelengths of  $\leq 10 \text{ km}$  can be observed concomitantly with the cloudiness transitions (Fig. 1, E and G).

Throughout the year in the southeast Atlantic, we also observed small-scale propagating cloud wave trains moving toward the south, southwest, southeast, and east. Figure 4 and movies S5 and S6 show an example when overlapping cloud wave trains moved southeastward, southwestward, and eastward coincident with westward-moving cloud

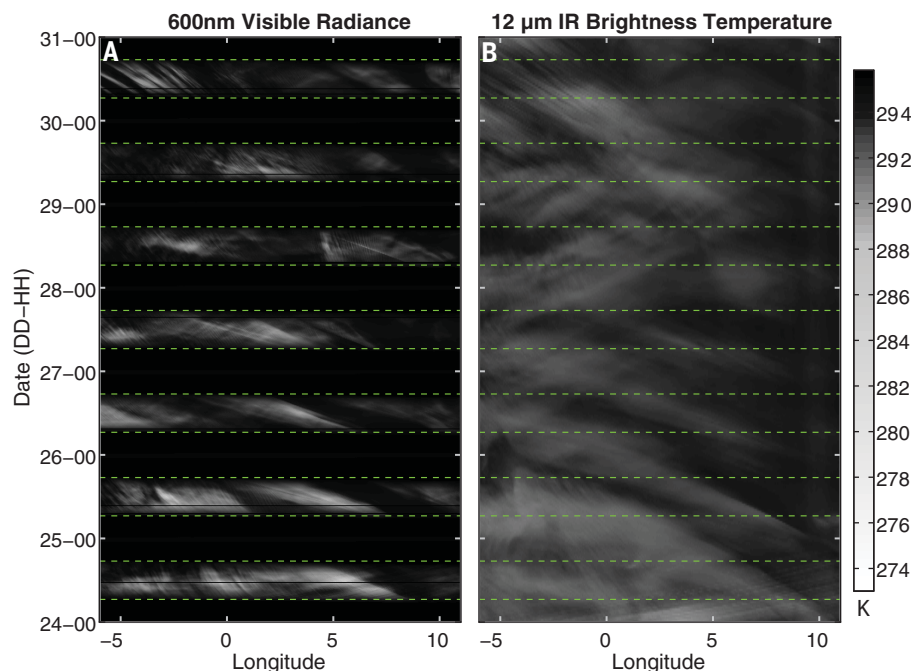
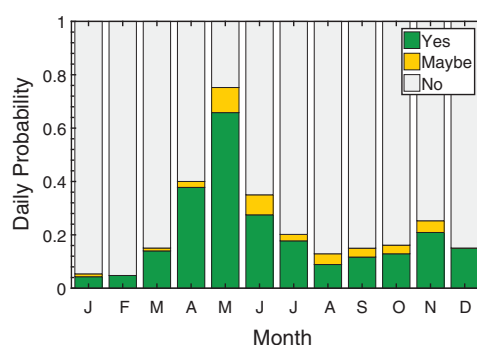


**Fig. 1. Examples of westward-moving cloudiness transitions in the southeast Atlantic off the coast of Africa in MODIS corrected reflectance imagery.** Consecutive pairs of Terra ( $\approx 10:30 \text{ a.m.}$ ) and Aqua ( $\approx 1:30 \text{ p.m.}$ ) images (A and B) on 25 May 2013 and (C and D) on

26 May 2014. (E to G) Boxed areas in (B) to (D) shown at greater magnification. Latitude and longitude grids are shown with dashed lines. Movie loop animations of the 26 May 2014 case are shown in movies S1 (reflectance) and S2 (IR).



**Fig. 2. Histogram of the daily probability of an abrupt cloudiness transition event occurring in a given month in the southeast Atlantic for the period from 8 May 2012 through 31 July 2017.** For example, the 0.658 daily probability value for the “Yes” events category for May means that on average 20.4 of the 31 days of May have westward-moving cloud erosion boundaries over the southeast Atlantic. The analysis was performed on the basis of pairs of Terra (morning) and Aqua (afternoon) visible corrected reflectance images. See methods for details.



**Fig. 3. Hovmöller plots illustrating cloud-clearing event timing and propagation speed.**

(A) Meteosat 600-nm visible imagery and (B) 12- $\mu$ m IR brightness temperature from 00 UTC on 24 May 2014 to 00 UTC on 31 May 2014. The dashed lines denote the approximate sunrise and sunset times. Data are averaged meridionally over a box from 6°W to 11°E longitude and 12.5°S to 17.5°S latitude. DD, day; HH, hour.

erosion. Cloud wave trains associated with deep convective latent heating were documented off the northwest coast of Australia during seasons when deep convection was common over the nearby continent (26). A potential source for southward-moving atmospheric gravity waves in the southeast Atlantic is latent heating from tropical deep convection (27) over western equatorial Africa and from the Atlantic Intertropical Convergence Zone. The eastward-moving wave trains in the southeast Atlantic may originate from the disturbed subtropical jet, similar to cloudiness perturbations seen in the southeast Pacific (23).

Cloud generation by atmospheric gravity waves is extensively documented in the literature, and the associated physical mechanisms are reasonably well understood (26, 28–32) (see supplement-

tary text for details). These studies show that when moisture and stability conditions are favorable, upward motions within the waves can yield propagating lines of clouds hundreds of kilometers in length moving on the order of 10 m s<sup>-1</sup>.

By contrast, cloud-eroding waves defy easy explanation. The upward and downward motion associated with atmospheric gravity wave passage should yield reversible changes. For a parcel with a given temperature and specific humidity, upward motion decreases the temperature, increases relative humidity, and, if saturation occurs, can yield cloud. Downward motion increases temperature and decreases relative humidity. If the moisture content of the air is unchanged, cloudiness would increase with upward motion, decrease with downward motion, and return to the prewave cloud state once the wave passes

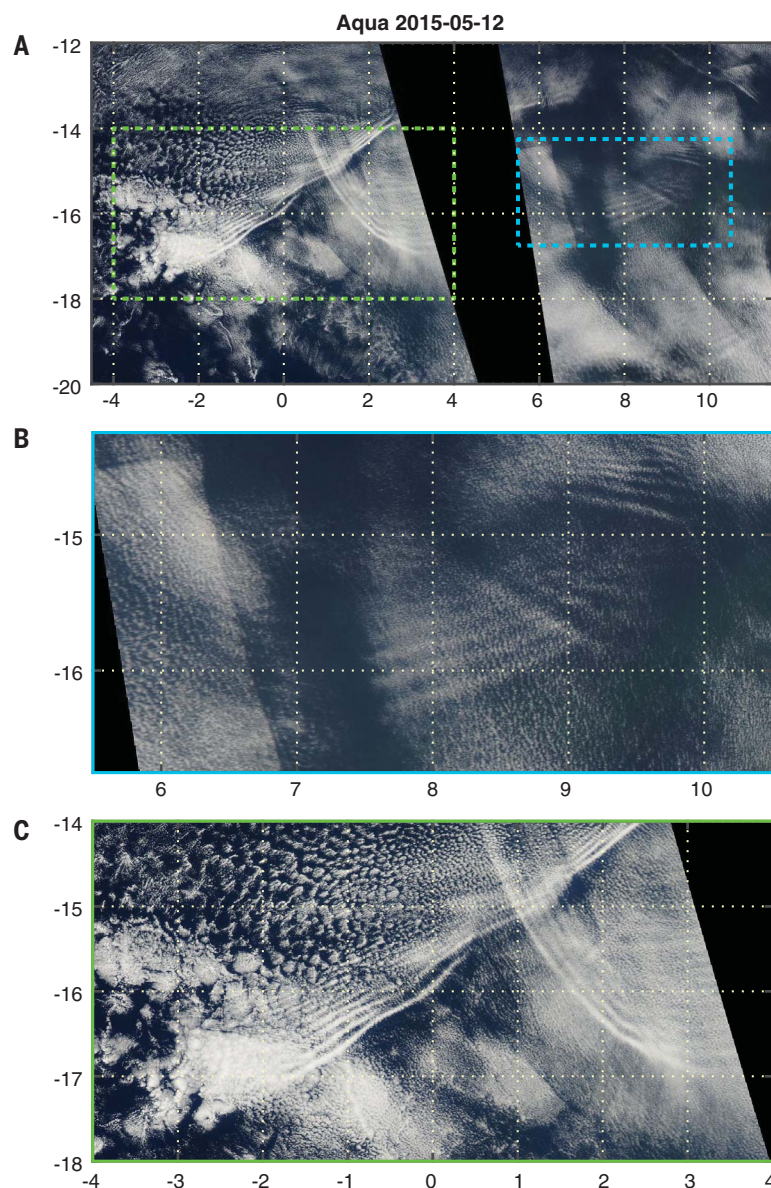
through. Yet observations show that cloud erodes along the boundaries in an irreversible manner that yields a reduced cloud fraction for hours (movies S2 and S4).

For atmospheric gravity wave-induced cloud clearing to occur, a wave must be excited and then must propagate through a layer of the atmosphere that includes the top of the marine boundary layer where the clouds reside. Further, the wave needs to be associated with a localized mechanism that can irreversibly reduce the cloud fraction on time scales of tens of minutes. Lastly, the cloud deck itself must be susceptible to this erosion by the atmospheric gravity wave.

Why abrupt cloud clearing over the subtropical southeast Atlantic is most frequent in May is related to some combination of environmental conditions favoring the presence of the above-mentioned factors in the region and the season of interest compared with other seasons and regions. MODIS aerosol optical depth measurements show that abrupt cloudiness transitions occur in a variety of aerosol conditions (fig. S1). Comparison of large-scale conditions derived from the Modern-Era Retrospective Analysis for Research and Applications (MERRA) for May, the month with the most cloud erosion boundaries, and for January, the month with the fewest, indicates weaker stability in May than in January (a median estimated inversion strength of 4.3 K versus 5.1 K) but stronger subsidence (700-hPa vertical pressure velocity of 0.049 Pa s<sup>-1</sup> versus 0.032 Pa s<sup>-1</sup>).

Our hypothesis is that westward-moving atmospheric gravity waves cause the abrupt cloudiness transitions in the southeast Atlantic and are triggered by the interaction of offshore flow—likely combining a nocturnal land breeze and downslope winds from the coastal highlands (33)—with the stable marine boundary layer in a manner similar to processes that generate cloud-forming atmospheric gravity waves (26, 28, 30, 31). For individual cloud erosion boundaries, we typically do not see evidence of slowing with increasing distance from the coast, as would be expected from a thinning, dissipating cold-air mass originating over land (Fig. 3). Among the three subtropical marine stratocumulus cloud decks in the northeast Pacific, southeast Pacific, and southeast Atlantic, the deck in the southeast Atlantic has the largest diurnal wind variability near the adjacent coast (34), which would be consistent with regular offshore flow events triggering gravity waves. Comparable westward-moving abrupt cloud-clearing events are not common in either the southeast Pacific or the northeast Pacific.

The removal of cloud over tens of minutes and the observation of bisected closed cellular cloud structures suggest a fast-acting mechanism. We hypothesize that cloud erosion is a consequence of rapid entrainment of warm and dry air from the free troposphere into the cloud layer by enhanced turbulence associated with a solitary wave train excited by offshore flow emanating from southwest Africa. Locally enhanced turbulent kinetic energy and mixing across the inversion



**Fig. 4. Examples of multiple small-scale cloud wave trains propagating southeastward, southwestward, and eastward over the southeast Atlantic.** MODIS corrected reflectance images centered off the west coast of Africa on 12 May 2015 during the afternoon Aqua overpass are shown. **(A)** Regional image with 2° grid lines. **(B and C)** Close-ups of waves overlaid with 1° grid lines corresponding to the inset boxes in (A). Movie loop animations of this case are shown in movies S5 (reflectance) and S6 (IR).

have been documented within a solitary wave packet in the central United States (35).

Targeted observations in the subtropical southeast Atlantic, including the use of dropsondes and airborne radar and lidar measurements, will be needed to resolve the mechanisms for rapid cloud erosion along cloud-eroding boundaries. The meteorological conditions associated with the formation and lack of formation of the cloudiness boundaries, as well as features of downslope winds and their interactions with the marine stable layer, are currently being investigated by combining observations, reanalysis, and mesoscale modeling.

Persistent, wide-area reductions in the cloud fraction over hundreds of kilometers associated with these westward-moving cloud erosion boundaries contribute to a lower March through May average cloud fraction in the subtropical southeast Atlantic than in any season in the subtropical southeast Pacific or the subtropical northeast Pacific (36). Mechanisms that yield substantial multiday variability in the marine stratocumulus cloud fraction are highly relevant to the climate system. Cloud-system resilience to external multiday perturbations and the frequency of these perturbations may be key factors

in governing low-cloud variability and potential climate sensitivity.

## REFERENCES AND NOTES

1. D. L. Hartmann, M. E. Ockert-Bell, M. L. Michelsen, *J. Clim.* **5**, 1281–1304 (1992).
2. C. S. Bretherton et al., *Bull. Am. Meteorol. Soc.* **85**, 967–978 (2004).
3. W. H. Schubert, J. S. Wakefield, E. J. Steiner, S. K. Cox, *J. Atmos. Sci.* **36**, 1286–1307 (1979).
4. B. Stevens, *Q. J. R. Meteorol. Soc.* **128**, 2663–2690 (2002).
5. B. Stevens, W. R. Cotton, G. Feingold, C.-H. Moeng, *J. Atmos. Sci.* **55**, 3616–3638 (1998).
6. A. S. Ackerman et al., *Mon. Weather Rev.* **137**, 1083–1110 (2009).
7. R. Wood et al., *Atmos. Chem. Phys.* **11**, 2341–2370 (2011).
8. I. Sandu, B. Stevens, *J. Atmos. Sci.* **68**, 1865–1881 (2011).
9. A. H. Berner, C. S. Bretherton, R. Wood, A. Muhlbauer, *Atmos. Chem. Phys.* **13**, 12549–12572 (2013).
10. S. P. de Szoeke, K. L. Verlinden, S. E. Yuter, D. B. Mechem, *J. Clim.* **29**, 6463–6481 (2016).
11. C. Brunet, R. Casotti, V. Vantrepotte, *J. Plankton Res.* **30**, 645–654 (2008).
12. C. M. Risien, D. B. Chelton, *J. Phys. Oceanogr.* **38**, 2379–2413 (2008).
13. S. A. Klein, *J. Clim.* **10**, 2018–2039 (1997).
14. M. A. Rozendaal, W. B. Rossow, *J. Atmos. Sci.* **60**, 711–728 (2003).
15. R. C. George, R. Wood, *Atmos. Chem. Phys.* **10**, 4047–4063 (2010).
16. N.-C. Lau, M. W. Crane, *Mon. Weather Rev.* **123**, 1984–2006 (1995).
17. S. A. Klein, C. Jakob, *Mon. Weather Rev.* **127**, 2514–2531 (1999).
18. G. L. Stephens, *J. Clim.* **18**, 237–273 (2005).
19. E. Crosbie et al., *J. Atmos. Sci.* **73**, 1083–1099 (2016).
20. R. Garreaud, R. Muñoz, *J. Clim.* **17**, 1699–1710 (2004).
21. C. W. O'Dell, F. J. Wentz, R. Bennartz, *J. Clim.* **21**, 1721–1739 (2008).
22. R. Wood, M. Köhler, R. Bennartz, C. O'Dell, *Q. J. R. Meteorol. Soc.* **135**, 1484–1493 (2009).
23. G. Allen et al., *Q. J. R. Meteorol. Soc.* **139**, 32–45 (2013).
24. P. J. Connolly et al., *Atmos. Chem. Phys.* **13**, 7133–7152 (2013).
25. S. A. Klein, D. L. Hartmann, *J. Clim.* **6**, 1587–1606 (1993).
26. C. E. Birch, M. J. Reeder, *Q. J. R. Meteorol. Soc.* **139**, 1311–1326 (2013).
27. B. E. Mapes, *J. Atmos. Sci.* **50**, 2026–2037 (1993).
28. R. H. Clarke, R. K. Smith, D. G. Reid, *Mon. Weather Rev.* **109**, 1726–1750 (1981).
29. D. R. Christie, K. J. Muirhead, R. H. Clarke, *Nature* **293**, 46–49 (1981).
30. F. Désalmand, A. Szantai, L. Picon, M. Desbois, *J. Geophys. Res.* **108** (D18), 8004 (2003).
31. J. C. B. da Silva, J. M. Magalhaes, *Int. J. Remote Sens.* **30**, 1161–1182 (2009).
32. P. A. Lutzak, *Weather Forecast.* **28**, 55–76 (2013).
33. T. Qian, C. C. Epifanio, F. Zhang, *J. Atmos. Sci.* **69**, 130–149 (2012).
34. S. T. Gille, *Geophys. Res. Lett.* **32**, L05605 (2005).
35. S. E. Koch et al., *Mon. Weather Rev.* **136**, 1373–1400 (2008).
36. C. D. Burleyson, S. E. Yuter, *J. Clim.* **28**, 2968–2985 (2015).
37. J. Hader, "Propagating, cloud-eroding boundaries in southeast Atlantic marine stratocumulus," thesis, North Carolina State University, Raleigh, NC (2016).
38. NASA, EOSDIS Worldview; <https://worldview.earthdata.nasa.gov/>.
39. EUMETSAT Product Navigator, <http://navigator.eumetsat.int/>.
40. J. Janowiak, B. Joyce, P. Xie, NCEP/CPC L3 half hourly 4km global (60S–60N) merged IR VI, A. Savtchenko, Ed. [Goddard Earth Sciences Data and Information Services Center (GES DISC), 2017]; <http://dx.doi.org/10.5067/P4HZB9N27EKU>.
41. S. E. Yuter, Southeast Atlantic westward-moving cloud erosion boundaries, Open Science Framework (2018); doi:10.17605/OSF.IO/KR4JS.

## ACKNOWLEDGMENTS

M. T. Bryant, E. S. Chan, L. Lovell, S. Rhodes, E. Scott, and L. Tomkins helped to analyze satellite data. **Funding:** This work was supported by U.S. Department of Energy grants DE-SC0006701, DE-SC0006736, and DE-SC0016522 and NSF AGS grants 1656237 and 1656314. **Author contributions:** All authors contributed to the formulation of the conceptual model, analysis design, interpretation of the data, and critical

revision of the article. J.D.H. performed the majority of the data acquisition and processing, with M.A.M. aiding in these efforts. S.E.Y. drafted the article, in part on the basis of J.D.H.'s thesis (37). **Competing interests:** None of the authors have competing interests. **Data and materials availability:** Satellite datasets used in this study are archived and accessible at <https://worldview.earthdata.nasa.gov/>, [http://navigator.eumetsat.int/discovery/Start/DirectSearch/DetailResult.do?f\(r0\)=EO:EUM:DAT:MSG:HRSEVIRI](http://navigator.eumetsat.int/discovery/Start/DirectSearch/DetailResult.do?f(r0)=EO:EUM:DAT:MSG:HRSEVIRI), and [https://disc.gsfc.nasa.gov/datasets/GPM\\_MERGIR\\_V1/summary?keywords=GPM\\_MERGIR\\_1](https://disc.gsfc.nasa.gov/datasets/GPM_MERGIR_V1/summary?keywords=GPM_MERGIR_1) (38–40). Dates and times for the 377 southeast Atlantic westward-moving cloud erosion boundaries are accessible at Open Science Framework (41).

SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/361/6403/697/suppl/DC1](http://www.sciencemag.org/content/361/6403/697/suppl/DC1)  
Materials and Methods

Supplementary Text  
Fig. S1  
Table S1  
References (42–56)  
Movies S1 to S6

24 November 2017; accepted 8 July 2018  
Published online 19 July 2018  
10.1126/science.aar5836



## RNA METABOLISM

# Mixed tailing by TENT4A and TENT4B shields mRNA from rapid deadenylation

Jaechul Lim<sup>1,2\*</sup>†, Dongwan Kim<sup>1,2\*</sup>, Young-suk Lee<sup>1,2\*</sup>, Minju Ha<sup>1,2</sup>, Mihye Lee<sup>1,2</sup>, Jinah Yeo<sup>1,2</sup>, Hyesik Chang<sup>1,2</sup>, Jaewon Song<sup>1,2</sup>, Kwangseog Ahn<sup>1,2</sup>, V. Narry Kim<sup>1,2†</sup>

RNA tails play integral roles in the regulation of messenger RNA (mRNA) translation and decay. Guanylation of the poly(A) tail was discovered recently, yet the enzymology and function remain obscure. Here we identify TENT4A (PAPD7) and TENT4B (PAPD5) as the enzymes responsible for mRNA guanylation. Purified TENT4 proteins generate a mixed poly(A) tail with intermittent non-adenosine residues, the most common of which is guanosine. A single guanosine residue is sufficient to impede the deadenylase CCR4-NOT complex, which trims the tail and exposes guanosine at the 3' end. Consistently, depletion of TENT4A and TENT4B leads to a decrease in mRNA half-life and abundance in cells. Thus, TENT4A and TENT4B produce a mixed tail that shields mRNA from rapid deadenylation. Our study unveils the role of mixed tailing and expands the complexity of posttranscriptional gene regulation.

**R**NA tailing (nontemplated nucleotide addition to the 3' end of RNA) is one of the most frequent RNA modifications (1, 2). We previously developed a method called TAIL-seq that reveals the 3' end sequences of the transcriptome (3). TAIL-seq allowed us to investigate the mechanism and function of non-canonical mRNA tailing, such as U tails. Uridylation, particularly oligo-uridylation, occurs selectively on the 3' ends of deadenylated mRNAs and promotes mRNA decay (4–6).

Notably, we discovered guanylation at the 3' termini of mRNAs in HeLa cells, expanding the complexity of mRNA tails (3). Unlike U tails, guanylated tails were found mainly among long A tails [ $\geq \sim 25$  nucleotides (nt)] and were detected in a form of single-nucleotide addition (Fig. 1A). Guanosine residues were located mainly at the 3' terminal or penultimate positions (Fig. 1B). Terminal uridine and cytidine residues were found on long A tails, albeit less frequently. Non-adenosine residues in other species—such as mouse, frog, fish, and fly—suggest a conserved modification mechanism (fig. S1A).

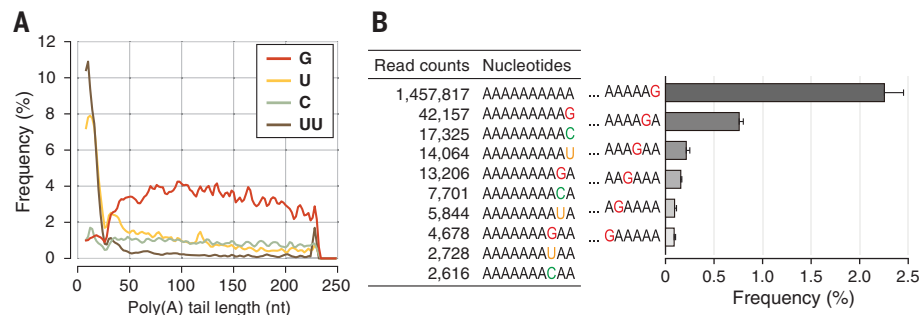
Tailing is catalyzed by a group of template-independent nucleotidyl transferases (7). Apart from the canonical poly(A) polymerases (PAPs) that cotranscriptionally generate mRNA poly(A) tails, vertebrates have multiple noncanonical PAPs with distinct substrate preferences (fig. S1B). Because some noncanonical PAPs catalyze uridylation instead of adenylation, noncanonical PAPs are also called terminal nucleotidyl-transferases (TENTs) (8, 9). To identify the enzyme(s) responsible for guanylation, we con-

ducted TAIL-seq after combinatorial knock-downs of the TENTs. Global guanylation frequency was reduced in TENT4A/4B-depleted (i.e., cells depleted of TENT4A and TENT4B), as well as in TENT1/4A/4B/6-depleted cells (fig. S1, C to E) (5). Individual knockdown of TENT4A or TENT4B did not affect guanylation (Fig. 2A), suggesting that these enzymes may act redundantly. Transcript-level analysis indicated that TENT4A and TENT4B may target a substantial fraction of mRNA species (656 out of 901, 72.8%) (Fig. 2B and table S1). A comparable result was obtained with a different set of small interfering RNAs (siRNAs) (fig. S1, E and F). Overexpression resulted in an increase of guanylation (Fig. 2C and fig. S1G), indicating that both TENT4A and TENT4B may catalyze guanylation. Of note, the previous annotation of TENT4A and TENT4B lacked conserved N-terminal extensions that are critical for their activity (fig. S2) (10, 11).

TENT4A (TUT5/PAPD7/hTRF4-1/POLS) and TENT4B (TUT3/PAPD5/hTRF4-2/GLD4) are homologs of yeast Trf4 (12). Trf4 was previously shown to catalyze guanylation, but the report has

been overlooked owing to the predominant adenylation activity of this protein (13). We found in vitro that TENT4B incorporates adenosine triphosphate (ATP) efficiently but also other nucleotides, with a preference for guanosine triphosphate (GTP) over uridine triphosphate (UTP) and cytidine triphosphate (CTP) (Fig. 2D). In vitro incorporation assay with TENT4B and nucleoside triphosphate (NTP) mixtures revealed that G residues were incorporated with a frequency of 15.5%, more than that of U (5.7%) and C (5.2%) residues (Fig. 2E). Consistently, uridylation and cytidylation frequencies decreased in TENT4A/B-depleted cells (fig. S1, H and I). Notably, non-adenosines were distributed without any positional preference at the 3' terminus (Fig. 2E, bottom right). Similar results were obtained with TENT4A (Fig. 2E, right), indicating that TENT4A and TENT4B generate a mixed tail with a preference for guanosine at irregular positions.

This in vitro result is seemingly inconsistent with our observation from TAIL-seq experiments, which showed a clear positional preference at the 3' end of poly(A) tails (Fig. 1B). To understand this discrepancy, we investigated whether additional factors are involved in tail remodeling. Deadenylases are 3'-5' exoribonucleases that favor poly(A) sequences (14–18). We hypothesized that deadenylase(s) may trim down the mixed tail and stall at non-A residues, thereby exposing non-A at the 3' terminus. The CCR4-NOT (CNOT) complex is the main cytoplasmic deadenylase complex and has two catalytic subunits: a Caf1 homolog (CAF1a/CNOT7 or CAF1b/CNOT8) and a Ccr4 homolog (CCR4a/CNOT6 or CCR4b/CNOT6L) (19, 20). When immunopurified CNOT7 proteins were incubated with synthetic poly(A) RNAs with intermittent Gs (A50G) (Fig. 3A), a ladderlike pattern was detected at the penultimate position (Fig. 3B, orange arrowheads). The same assay with catalytic mutants showed no ladderlike accumulation (fig. S3A). CNOT6L sheared the adenosine sequences but stalled at each G residue, yielding G at the 3' terminus (Fig. 3C, red arrowheads). Similarly, C and U residues were able to expose the non-adenosine residues (fig. S3B). Thus, the CNOT complex trims the mixed tail to generate a poly(A) tail with a terminal or penultimate non-adenosine residue.



**Fig. 1. Pervasive 3' end guanylation of the poly(A) tail.** (A) Omnipresent 3' end guanylation across poly(A) tails of length  $\geq 25$  nt. (B) Abundant guanosine residues at the 3' end of the poly(A) tail. (Left) The most abundant 3' 10-nucleotide sequences. (Right) Terminal and internal guanylation frequencies ( $n = 4$  TAIL-seq experiments). Error bars represent SEM.

<sup>1</sup>Center for RNA Research, Institute for Basic Science, Seoul 151-742, Republic of Korea. <sup>2</sup>School of Biological Sciences, Seoul National University, Seoul 151-742, Republic of Korea. \*These authors contributed equally to this work.

†Present address: Department of Immunobiology, Yale University School of Medicine, New Haven, CT 06510, USA.

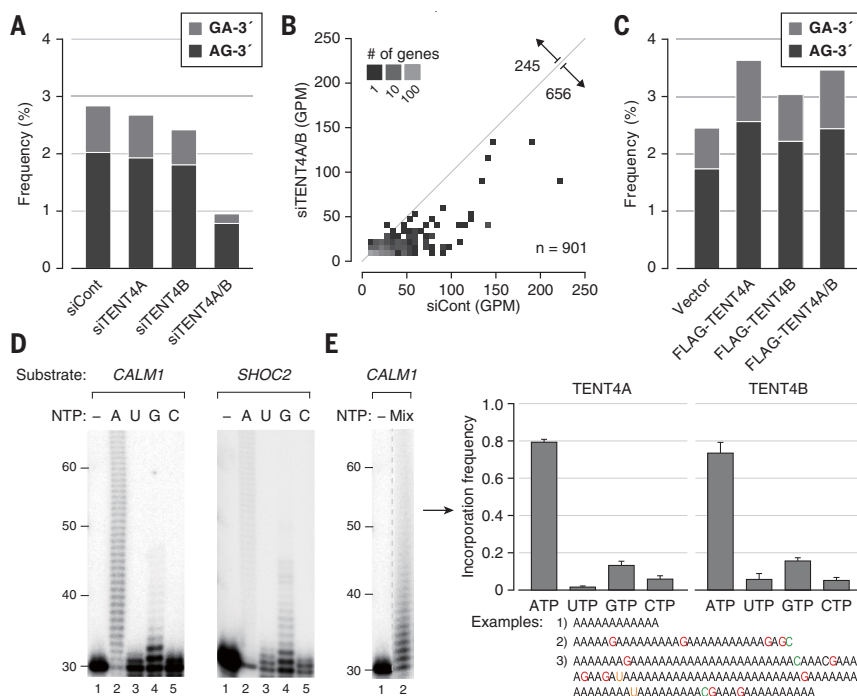
‡Corresponding author. Email: narrykim@snu.ac.kr

These results also indicate that even a single non-adenosine residue can counteract deadenylation and stabilize mRNA. Indeed, *in vitro* deadenylation was substantially delayed in the presence of a single terminal G (Fig. 3D, left, and fig. S3, C and D). Likewise, 3' penultimate or internal G was sufficient to impede deadenylation (Fig. 3D, right, and fig. S3, D and E). Consistently, the poly(A) tail length of mRNA was reduced in TENT4A/B-depleted cells (fig. S4A). Transcripts with statistically high guanylation were affected more strongly than those with low guanylation (fig. S4B), supporting our hypothesis that guanylation delays deadenylation.

To further examine whether TENT4A and TENT4B have a positive impact on mRNA stability, we performed RNA sequencing in TENT4A/B-depleted cells. A modest transcriptome-wide decrease in mRNA levels was observed (Fig. 4A and fig. S4C and table S2). Transcripts with statistically high guanylation were reduced compared with those that have low guanylation (Fig. 4B and table S3). Another siRNA set produced similar results (fig. S4D). mRNAs with frequent mixed tailing (including G, U, and C) were also destabilized (fig. S4E), consistent with our *in vitro* deadenylation experiments with non-adenosine residues (fig. S3B). Highly guanylated genes were destabilized in TENT4A/B-depleted cells after actinomycin D treatment (Fig. 4C), indicating that TENT4A/B-mediated mixed tailing may stabilize mRNAs.

Notably, transcripts encoding secreted proteins are frequently guanylated, suggesting an association between mixed tailing and endoplasmic reticulum-mediated translation (table S3). Conversely, many genes involved in protein synthesis (such as ribosomal proteins and translation factors) are not guanylated frequently, indicating that their mRNAs are immune to TENT4A/B-mediated regulation. Highly guanylated mRNAs have longer 3' untranslated regions (3'UTRs) (fig. S4F), suggesting that the specificity of TENT4A and TENT4B may reside in the 3'UTR. Our motif search did not reveal a particular motif enriched in the UTR, possibly because multiple elements may contribute in combination. TENT4B was previously reported to adenylate and destabilize miR-21 (27), but neither an increase in miR-21 level nor a significant change in microRNA (miRNA) targets was found in TENT4A/B-depleted cells (fig. S5). TENT4A/B homolog Trf4 is a component of the TRAMP (Trf4/Air1/2/Mtr4) complex (22–26), but guanylation was not reduced in MTR4-depleted cells (fig. S6). The results suggest that TENT4A and TENT4B regulate mRNA levels in a miRNA- or TRAMP complex-independent manner.

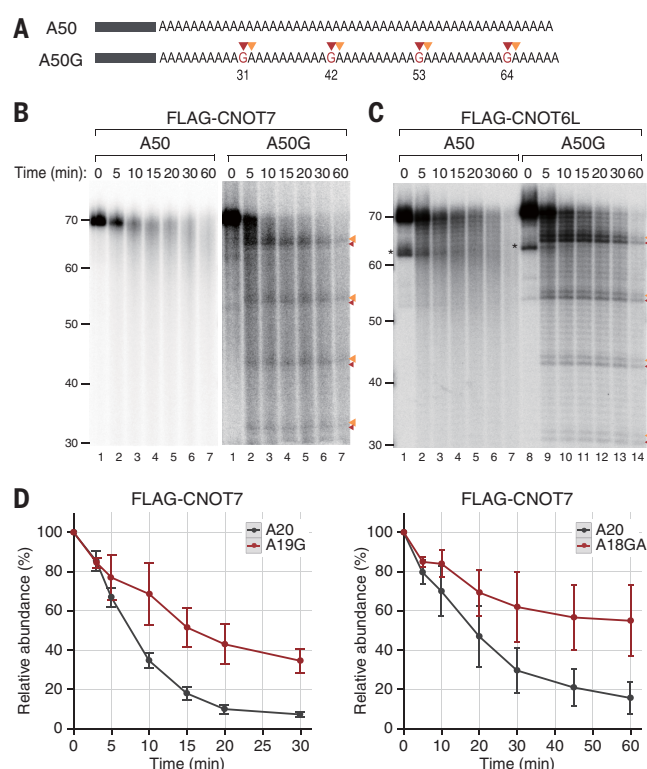
To measure the impact of mixed tailing on mRNA stability, we performed reporter tethering experiments (Fig. 4D, left). The canonical PAP that selectively incorporates adenosine was used as a negative control for the intrinsic adenylation activity of TENT4A (fig. S7A). Protein and mRNA levels were up-regulated when the reporter was tethered to wild-type TENT4A but not to its mutant (Fig. 4D, right, and fig. S7B). Similar effects

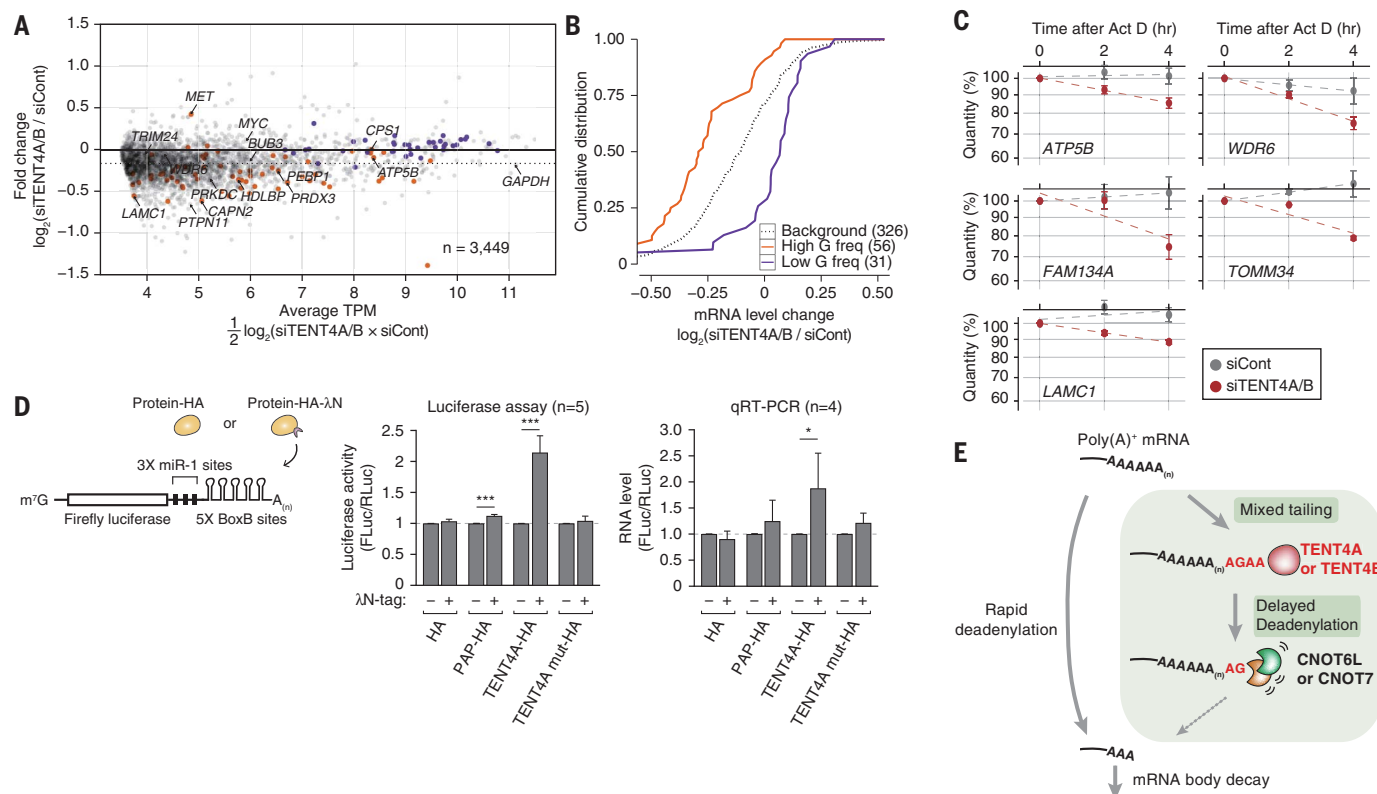


**Fig. 2. TENT4A and TENT4B are responsible enzymes for guanylation.** (A) Decrease in guanylation after knockdown of TENT4A and TENT4B. (B) Transcriptome-wide decrease in guanylation after TENT4A and TENT4B knockdown. GPM (guanylated tags per million) was calculated for each transcript. Transcript-level GPMs >3 before or after knockdown of TENT4A and TENT4B were illustrated as a two-dimensional density plot. Cont, control. (C) Increase in guanylation upon overexpression of TENT4A and TENT4B. (D) In vitro incorporation assay with individual NTP and FLAG-TENT4B. (E) In vitro incorporation assays with mixed NTPs showed intermittent guanylation. (Left) In vitro assays with FLAG-TENT4B. “Mix” refers to an NTP mixture (ATP:UTP:GTP:CTP) of 1:1:1:1 ratio. A dashed line shows discontinuous lanes from the same gel. (Right) Extended tails by TENT4A and TENT4B were sequenced by Sanger sequencing ( $n = 28$  and 15, respectively). Error bars indicate SEM.

### Fig. 3. Stalled deadenylation of CNOT7 and CNOT6L leads to the accumulation of guanosine at the 3' end.

(A) Illustration of RNA substrates. Guanosine residues are colored in red, and distances (in nucleotides) from the 5' end are labeled. Red and orange arrowheads denote the 3' end of G and the subsequent A (GA), respectively. (B) In vitro deadenylation assay with FLAG-CNOT7. (C) In vitro deadenylation assay with FLAG-CNOT6L. Asterisks mark chemical artifacts from oligonucleotide synthesis. (D) In vitro deadenylation assays with A20, A19G (left), or A18GA (right) substrates and FLAG-CNOT7. The relative abundance of intact RNAs (of length 20 nt) was compared by mean estimation and SEM error bars ( $n = 3$ ).





**Fig. 4. Mixed tailing by TENT4A and TENT4B stabilizes mRNA.**

(A) Decrease in steady-state mRNA levels after knockdown of TENT4A and TENT4B knockdown ( $n = 2$ ). The dashed horizontal line indicates the median fold change. Labeled mRNAs were validated by quantitative reverse transcription polymerase chain reaction (qRT-PCR) (fig. S4C). TPM, transcripts per million. (B) Upon knockdown of TENT4A and TENT4B, mRNA levels of highly guanylated genes ( $n = 56$ ;  $P < 0.01$ ) decreased more than those of genes with low guanylation ( $n = 31$ ;  $P < 0.01$ ). “Background” refers

to genes with more than five G and GA counts. Of note, this background cumulative distribution is similar to that of all genes. (C) Measurement of half-lives by qRT-PCR ( $n = 3$ ). GAPDH (glyceraldehyde phosphate dehydrogenase) was used for normalization. Act D, actinomycin D. (D) Tethering experiments. (Left) Experimental scheme. (Right) Reporter protein and mRNA levels were determined by luciferase assays and qRT-PCR, respectively. \* $P < 0.05$ , \*\*\* $P < 0.0001$ ; two-sided  $t$  test. Error bars indicate SEM. (E) Proposed model.

were found with TENT4B-tethered reporters (fig. S7, C and D). Thus, the non-adenosine incorporation of TENT4A and TENT4B may be important for mRNA stabilization, although this experiment does not rule out the possible contribution of adenylation. Of note, the same enzymology and function were observed in human foreskin fibroblast (HFF) cells (fig. S8). Growth defects were found in TENT4A/B-depleted HeLa and HFF cells (fig. S9), hinting at the physiological importance of TENT4/B-dependent mRNA tailing (27–29).

The mRNA tail is thought to be a pure stretch of As, with little informational content except for its length. In this study, we demonstrate that TENT4A and TENT4B generate specialized tails that protect mRNAs from active deadenylation (Fig. 4E). The protective effect may be attributed to both adenylation and mixed tailing. We focused on the impact of non-adenosines added by TENT4A and TENT4B and discovered the considerable stalling effect. The wide range of mRNA destabilization in TENT4A/B-depleted cells opens the possibility of multiple mixed tailing events during the mRNA life span. The “not-so-pure” poly(A) tail is constantly remodeled by tailing enzymes and exoribonucleases and acts as an

additional layer for posttranscriptional regulation, in effect changing the way we think about mRNA tails.

#### REFERENCES AND NOTES

- C. J. Norbury, *Nat. Rev. Mol. Cell Biol.* **14**, 643–653 (2013).
- M. Lee, B. Kim, V. N. Kim, *Cell* **158**, 980–987 (2014).
- H. Chang, J. Lim, M. Ha, V. N. Kim, *Mol. Cell* **53**, 1044–1052 (2014).
- O. S. Rissland, C. J. Norbury, *Nat. Struct. Mol. Biol.* **16**, 616–623 (2009).
- J. Lim et al., *Cell* **159**, 1365–1376 (2014).
- H. Chang et al., *Mol. Cell* **70**, 72–82.e7 (2018).
- L. Aravind, E. V. Koonin, *Nucleic Acids Res.* **27**, 1609–1618 (1999).
- T. E. Mullen, W. F. Marzluff, *Genes Dev.* **22**, 50–65 (2008).
- C. J. Willusz, J. Willusz, *Genes Dev.* **22**, 1–7 (2008).
- C. Rammelt, B. Bilen, M. Zavolan, W. Keller, *RNA* **17**, 1737–1746 (2011).
- K. Ogami, R. Cho, S. Hoshino, *Biochem. Biophys. Res. Commun.* **432**, 135–140 (2013).
- J. Houseley, D. Tollervey, *Cell* **136**, 763–776 (2009).
- J. LaCava et al., *Cell* **121**, 713–724 (2005).
- P. Viswanathan, J. Chen, Y. C. Chiang, C. L. Denis, *J. Biol. Chem.* **278**, 14949–14955 (2003).
- C. Bianchini, F. Mauviel, S. Sentis, B. Séraphin, L. Corbo, *RNA* **11**, 487–494 (2005).
- N. Henriksson, P. Nilsson, M. Wu, H. Song, A. Virtanen, *J. Biol. Chem.* **285**, 163–170 (2010).

- H. Wang et al., *EMBO J.* **29**, 2566–2576 (2010).
- S. Niinuma, T. Fukaya, Y. Tomari, *RNA* **22**, 1550–1559 (2016).
- Y. B. Yan, *RNA* **5**, 421–443 (2014).
- H. Yi et al., *Mol. Cell* **70**, 1081–1088.e5 (2018).
- J. Boele et al., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 11467–11472 (2014).
- H. Berndt et al., *RNA* **18**, 958–972 (2012).
- C. Kilchert, S. Wittmann, L. Vasiljeva, *Nat. Rev. Mol. Cell Biol.* **17**, 227–239 (2016).
- B. Boyraz et al., *J. Clin. Invest.* **126**, 3377–3382 (2016).
- S. Shukla, J. C. Schmidt, K. C. Goldfarb, T. R. Cech, R. Parker, *Nat. Struct. Mol. Biol.* **23**, 286–292 (2016).
- H. Sudo, A. Nozaki, H. Uno, Y. Ishida, M. Nagahama, *FEBS Lett.* **590**, 2963–2972 (2016).
- M. Schmid, B. Küchler, C. R. Eckmann, *Genes Dev.* **23**, 824–836 (2009).
- D. M. Burns, A. D’Ambrogio, S. Nottrott, J. D. Richter, *Nature* **473**, 105–108 (2011).
- J. Shin, K. Y. Paek, M. Ivshina, E. E. Stackpole, J. D. Richter, *Nucleic Acids Res.* **45**, 6793–6804 (2017).

#### ACKNOWLEDGMENTS

We thank members of our laboratory for discussion and technical help, especially H. Yi for CNOT7/6L constructs and E. Kim for plasmid cloning. **Funding:** This work was supported by IBS-R008-D1 of Institute for Basic Science from the Ministry of Science, ICT and Future Planning of Korea (J.L., D.K., Y.-s.L., M.H., M.L., J.Y., H.C., J.S., K.A., and V.N.K.) and BK21 Research Fellowships from the Ministry of Education of Korea (D.K. and J.Y.). **Author contributions:** J.L., D.K., and V.N.K. designed experiments.



J.L., D.K., and M.H. performed biochemical and cell biological experiments. M.L. and J.Y. performed TAIL-seq on model organisms. Y.-s.L. and J.L. carried out computational analyses. J.S. and K.A. provided primary HFF cells. J.L., Y.-s.L., and V.N.K. wrote the manuscript. Y.-s.L. and H.C. provided statistical and technical support. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Sequenced reads have been deposited in the NCBI

Gene Expression Omnibus (GEO) database (accession number GSE116355).

**SUPPLEMENTARY MATERIALS**

[www.sciencemag.org/content/361/6403/701/suppl/DC1](http://www.sciencemag.org/content/361/6403/701/suppl/DC1)

Materials and Methods

Figs. S1 to S9

Tables S1 to S5

References (30–33)

13 December 2016; resubmitted 17 March 2018

Accepted 10 July 2018

Published online 19 July 2018

10.1126/science.aam5794

## IMMUNOLOGY

# DNA-induced liquid phase condensation of cGAS activates innate immune signaling

Mingjian Du<sup>1,2</sup> and Zhijian J. Chen<sup>1,2,3\*</sup>

The binding of DNA to cyclic GMP–AMP synthase (cGAS) leads to the production of the secondary messenger cyclic GMP–AMP (cGAMP), which activates innate immune responses. We have shown that DNA binding to cGAS robustly induced the formation of liquidlike droplets in which cGAS was activated. The disordered and positively charged cGAS N terminus enhanced cGAS–DNA phase separation by increasing the valencies of DNA binding. Long DNA was more efficient in promoting cGAS liquid phase separation and cGAS enzyme activity than short DNA. Moreover, free zinc ions enhanced cGAS enzyme activity both in vitro and in cells by promoting cGAS–DNA phase separation. These results demonstrated that the DNA-induced phase transition of cGAS promotes cGAMP production and innate immune signaling.

Cyclic GMP–AMP synthase (cGAS) is a DNA-sensing enzyme that catalyzes the conversion of GTP and ATP to cyclic GMP–AMP (cGAMP), which activates the adaptor protein STING (1, 2). This, in turn, induces type I interferons and other cytokines (3–5). DNA arising in the cytoplasm activates cGAS and drives the formation of cytoplasmic foci containing cGAS and DNA (1). However, the molecular mechanism and functional effect of such cGAS foci are poorly understood. cGAS contains a disordered and positively charged N terminus and a structured C terminus harboring a nucleotidyltransferase domain (core cGAS). Both the N and C termini of cGAS bind to DNA irrespective of DNA sequence (6–11). We hypothesized that such multivalent interactions could lead to the formation of large membraneless protein foci through liquid phase separation (Fig. 1A). This physicochemical process has emerged as a key mechanism underlying the formation of cellular bodies such as P granules and nucleoli (12–20).

To test whether DNA binding induces phase separation of cGAS in vitro, we incubated fluorescently labeled cGAS protein (fig. S1A) with double-stranded DNA (dsDNA) oligonucleotides [100 base pairs (bp)] (table S1 and methods). Upon mixing, cGAS and DNA formed micrometer-sized liquid droplets within 2 min (Fig. 1B and movie S1). Small liquid droplets fused into larger ones (Fig. 1C), accompanied by increased fluorescence intensity and larger equivalent diameter (Fig. 1, D and E, and movie S1). Fluorescence recovery after photobleaching (FRAP) experiments showed that when bleaching was performed 30 min after

the initiation of phase separation, the fluorescence of cGAS or DNA was efficiently recovered in a temperature-dependent manner (Fig. 1F and fig. S1, B to D). In contrast, when bleaching occurred 2 hours after mixing of cGAS and DNA, the fluorescence recovery was much slower (Fig. 1F and fig. S1D). Thus, cGAS and DNA molecules within the liquid droplets are mobile and exhibit dynamic internal rearrangement in the early phase but gradually undergo a liquid-to-solid transition and mature into a gel-like state (Fig. 1F) (19, 21).

cGAS and DNA formed liquid droplets when the concentration of each exceeded 30 nM (Fig. 1G) in a buffer mimicking physiological ion concentrations and the composition of the cytoplasm (physiological buffer) (methods). Measurements of cGAMP production by cGAS revealed that the specific activity of cGAS substantially increased at concentrations above 30 nM (Fig. 1H and methods). The in vitro cGAS–DNA phase transition was weakened by increasing salt (NaCl) concentrations (fig. S2A), suggesting that ionic interactions between cGAS and DNA were important for the phase transition. A 45-bp dsDNA commonly used for the stimulation of the cGAS pathway [immune stimulatory DNA (ISD)] also robustly induced cGAS liquid droplet formation (fig. S2, B to D). This effect was abolished by treatment with Benzonase, which degrades DNA (fig. S2E). cGAS also formed liquid droplets with 45-bp double-stranded RNA (fig. S2F), but RNA did not activate cGAS to produce cGAMP (fig. S2G). These results are consistent with previous structural studies indicating that DNA but not RNA binding induces a conformational change that activates cGAS (7). Thus, liquid phase separation is insufficient for cGAS activation in the absence of the correct conformational change induced by DNA.

The cGAS–DNA phase separation was unaffected by the addition of ATP, GTP, or a combination thereof (fig. S3A). Moreover, ATP or GTP could be partitioned into and enriched in

the cGAS–DNA liquid droplets (fig. S3, B and C). FRAP experiments showed that ATP was rapidly exchanged into and out of the cGAS–DNA droplets (fig. S3D).

We next examined the formation of cGAS–DNA foci within cells. In the human fibroblast cell line BJ-5ta stably expressing a Halo-tagged cGAS, cGAS formed puncta with fluorescein-labeled ISD in the cytoplasm (Fig. 2A and fig. S4A). To confirm that cGAS formed large granules with DNA in the cytoplasm, we used cGAS-deficient mouse embryonic fibroblast (MEF) cells reconstituted with green fluorescent protein (GFP)–cGAS, transfected the cells with Cy5-ISD, and then permeabilized the cells with saponin (22). cGAS formed puncta with ISD and remained in the cytoplasm after saponin treatment (Fig. 2B). The cGAS–DNA foci exhibited liquidlike properties, as demonstrated by the ability of two foci to fuse with each other (Fig. 2C). Furthermore, upon photobleaching, cGAS in the foci displayed near-complete fluorescence recovery within 120 s (Fig. 2, D and E), indicating that cGAS exhibits dynamic liquidlike behavior within cellular granules.

To determine the functional consequence of cGAS liquid droplet formation, we transfected HeLa cells with herring testis DNA (HT-DNA) and measured cGAS activity in subcellular fractions (fig. S4B). Most cGAS activity was present in the pellet obtained after centrifugation at  $2,000 \times g$  (designated P2), which contains mainly the nuclei and heavy particles (Fig. 2F). Further separation of the P2 fraction by iodixanol (OptiPrep) density gradient ultracentrifugation revealed two distinct pools of cGAS activity. The first pool was in very heavy fractions (20 to 25% iodixanol) (Fig. 2G), which were separated from the nuclei (27.5 to 30% iodixanol). Thus, cGAS appeared to form heavy particles with transfected DNA that were distinct from cellular organelles and vesicles and that contained active cGAS. The second pool of cGAS activity was in fractions containing the nuclei. However, it remains to be determined whether this activity came from cGAS within the nuclei or some cGAS particles that cosedimented with the nuclei. Similar results were also obtained with human monocytic THP-1 cells (fig. S4, C and D).

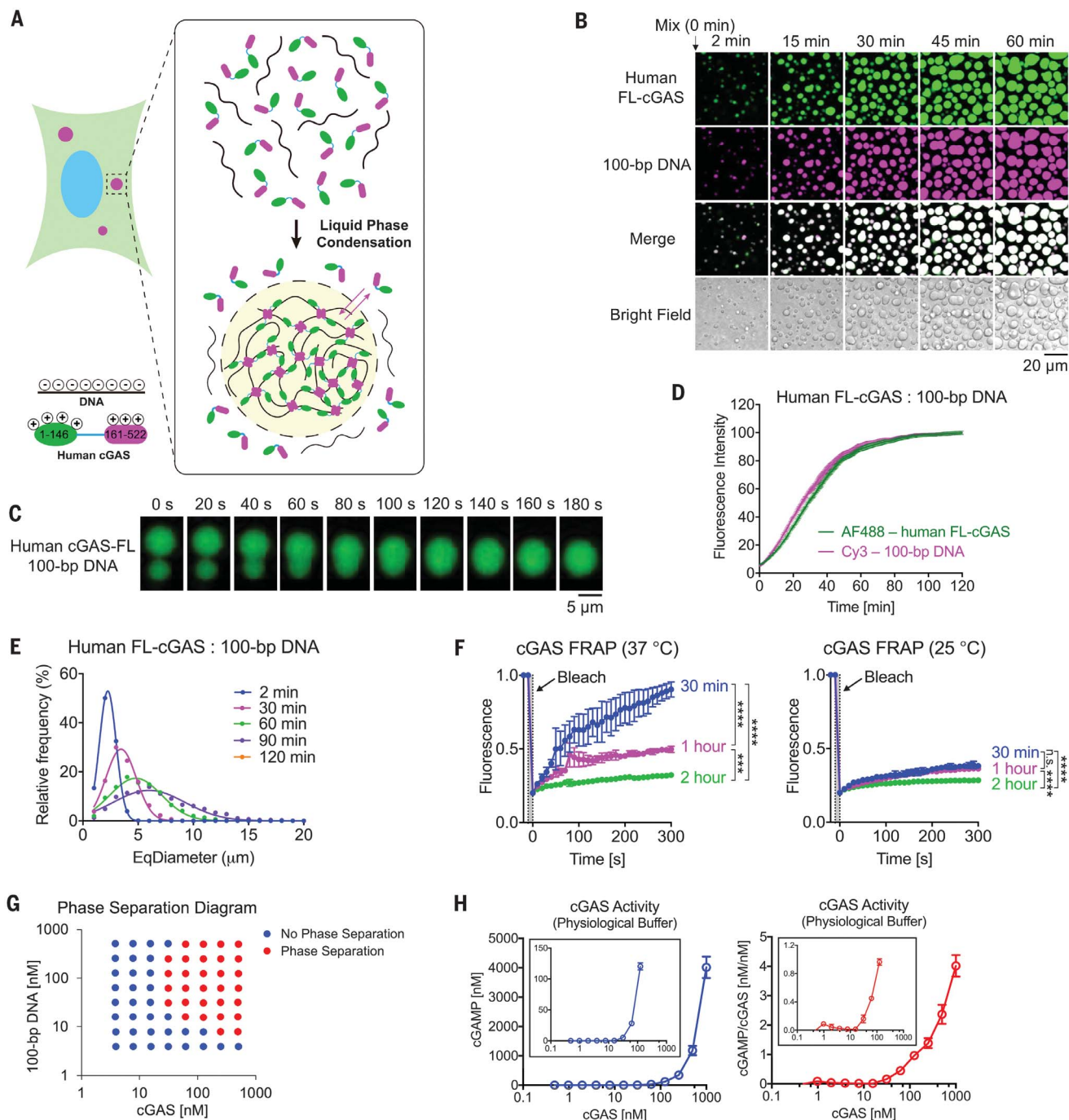
Multivalent interactions drive liquid phase separation (16). Long DNA has more binding sites (higher valency) for cGAS than short DNA, and full-length cGAS (FL-cGAS) has higher valency for DNA than core cGAS (Fig. 3A) (6, 7, 11). To test whether cGAS–DNA liquid phase separation is driven by the valency of cGAS and DNA interactions, we incubated FL-cGAS and N-terminally truncated cGAS (AN146-cGAS, where the numeral indicates the number of amino acids deleted) with DNA of different lengths in the physiological buffer (15 mM NaCl and 135 mM KCl) or a buffer containing 300 mM NaCl. Human FL-cGAS formed more numerous and larger liquid droplets with longer DNA (Fig. 3B). Both human and mouse FL-cGASs exhibited stronger phase separation than N-terminally truncated cGAS with DNA of the same length either in the physiological

<sup>1</sup>Department of Molecular Biology, University of Texas Southwestern Medical Center, Dallas, TX 75390-9148, USA.

<sup>2</sup>Center for Inflammation Research, University of Texas Southwestern Medical Center, Dallas, TX 75390-9148, USA.

<sup>3</sup>Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, TX 75390-9148, USA.

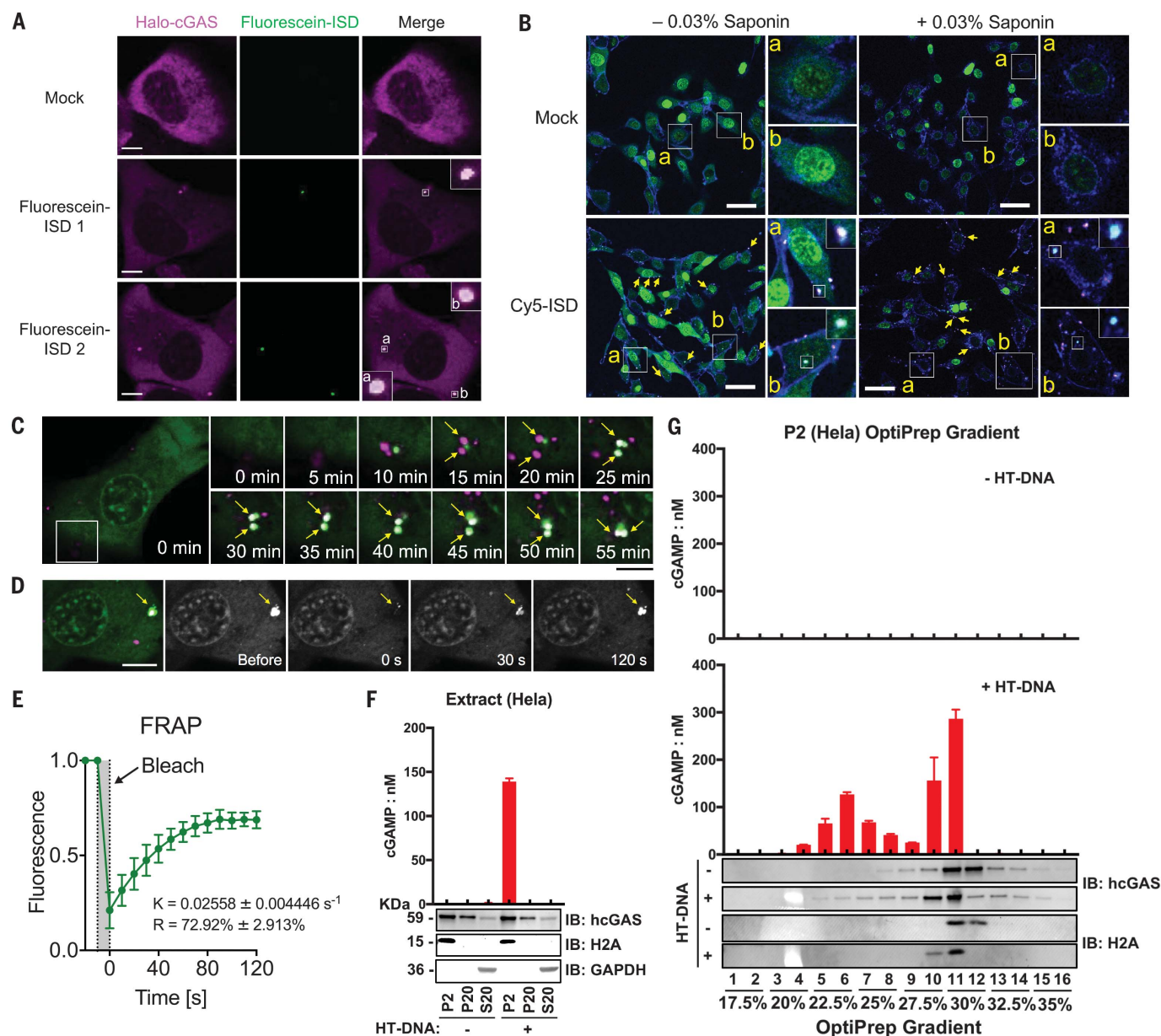
\*Corresponding author. Email: zhijian.chen@utsouthwestern.edu



**Fig. 1. DNA binding to cGAS induces the formation of liquidlike droplets in which cGAS is activated.** (A) Schematic of hypothetical cGAS-DNA interactions that drive liquid phase condensation. (B) Time-lapse imaging of cGAS-DNA phase separation. Liquid droplets formed after mixing of 10  $\mu\text{M}$  human FL-cGAS (3% was labeled with Alexa 488) with 10  $\mu\text{M}$  100-bp DNA (2% was labeled with Cy3) and matured over 60 min. The images shown are representative of all fields in the well. (C) Time-lapse micrographs of merging droplets that formed as described in the legend to (B). Data are representative of at least 10 merging droplets. (D) Fluorescence intensities of cGAS-DNA liquid droplets forming over the time course of 120 min. Data were normalized to 100% by maximal fluorescence intensity. Values shown are means  $\pm$  SD.  $n = 4$  images. AF488, Alexa Fluor 488. (E) EqDiameter (the diameter of a circle with the same area as the measured object) frequency distribution of cGAS-DNA liquid

droplets formed at the indicated time points. cGAS, 5  $\mu\text{M}$ ; DNA, 5  $\mu\text{M}$ . (F) FRAP of cGAS-DNA liquid droplets. Bleaching was performed at the indicated time points after cGAS (10  $\mu\text{M}$ ) and DNA (10  $\mu\text{M}$ ) were mixed, and the recovery was allowed to occur at 25° or 37°C. Time 0 indicates the start of recovery after photobleaching. Shown are the means  $\pm$  SD.  $n = 3$  liquid droplets. n.s., not significant ( $P > 0.0332$ ); \*\*\* $P < 0.0002$ ; \*\*\*\* $P < 0.0001$  [one-way analysis of variance (ANOVA)]. (G) Phase separation diagram of human FL-cGAS and 100-bp DNA at the indicated concentrations. (H) (Left) cGAMP production by the indicated concentrations of cGAS in the presence of ATP, GTP, and HT-DNA. cGAMP production at low cGAS concentrations is shown in the inset. (Right) Normalized cGAMP production divided by cGAS concentrations. Shown are the means  $\pm$  SD.  $n = 3$  assays. Data are representative of at least three independent experiments unless indicated to be otherwise.





**Fig. 2. DNA-induced liquid phase separation of cGAS in cells.**

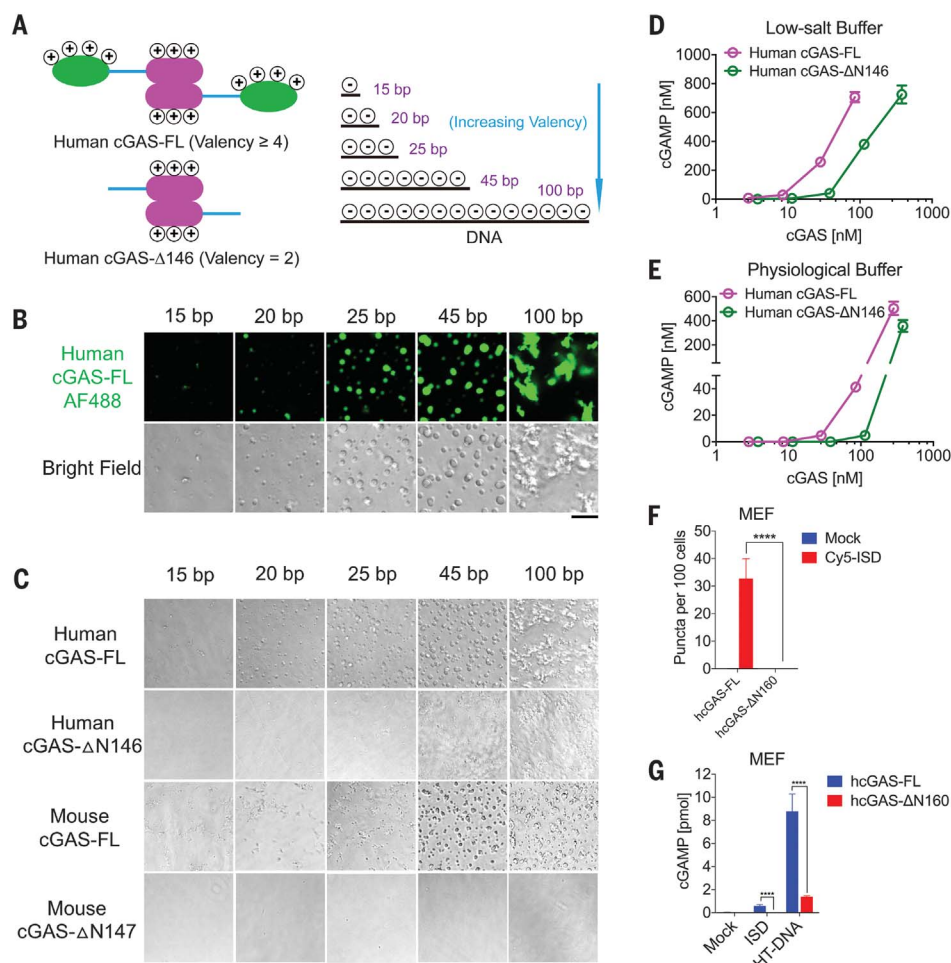
(A) Representative live-cell images of cGAS-DNA puncta formation after BJ cells stably expressing Halo-cGAS, which was covalently labeled with tetramethyl rhodamine, were transfected with fluorescein-ISD. Insets are zoomed images showing cGAS-DNA puncta. Scale bars, 10  $\mu\text{m}$ . These images are representative of at least 10 cells. (B) MEF cells stably expressing GFP-cGAS were transfected with Cy5-ISD for 4 hours, after which they were permeabilized with saponin and analyzed by fluorescence microscopy. Shown in blue is the plasma membrane marker wheat germ agglutinin. Arrows indicate puncta. Scale bars, 50  $\mu\text{m}$ . These images are representative of at least five fields examined. (C) Time-lapse micrographs of cGAS (green) and DNA (magenta) puncta formation and fusion (time 0 represents 30 min after transfection with Cy5-45-bp ISD). Arrows indicate puncta. Scale bar, 15  $\mu\text{m}$ . The fusion events existed in all eight fields examined. (D) Representative micrographs of cGAS-DNA puncta before and after photobleaching (arrow, bleach site). Scale bar, 15  $\mu\text{m}$ . These images are representative of at least

three cells in which the cGAS-DNA puncta were photobleached.

(E) Quantification of FRAP of cGAS-DNA puncta over a 120-s time course.  $K$ , exponential constant;  $R$ , normalized plateau after fluorescence recovery. Shown are means  $\pm$  SD.  $n = 3$  cGAS-DNA puncta.

(F) Subcellular fractionation of cGAS activity in cells transfected with DNA. HeLa cells transfected with HT-DNA or untransfected HeLa cells were fractionated by differential centrifugation as depicted in fig. S4B. Fractions were incubated with ATP and GTP, after which cGAMP was measured. Fractions were also analyzed by immunoblotting (IB) with antibodies against histone H2A (nuclear marker), glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (cytoplasmic marker), or human cGAS (hcGAS).

(G) The P2 fractions from (F) were further separated by OptiPrep gradient ultracentrifugation, and cGAS activity in different fractions was measured as for (F). Fractions from cells not transfected with DNA had no cGAS activity (upper panel). Error bars in (F) and (G) represent the variation range of duplicate assays. Data are representative of at least three independent experiments.



**Fig. 3. Multivalent interactions drive cGAS-DNA condensation and promote cGAS activation.**

(A) Schematic of hypothetical cGAS and DNA valencies. (B) Representative images of phase separation by mixing of cGAS (10  $\mu$ M) with dsDNA of different lengths (10  $\mu$ M) in physiological buffer. Scale bar, 10  $\mu$ m. (C) Bright-field images of phase separation by mixing of DNA of different lengths with full-length or N-terminally truncated human or mouse cGAS as indicated. Scale bar, 20  $\mu$ m. The images shown in (B) and (C) are representative of all fields in the wells. (D and E) cGAMP production by different concentrations of recombinant human FL-cGAS or N-terminally truncated cGAS in low-salt buffer (D) or physiological buffer (E). Shown are the means  $\pm$  SD.  $n = 3$  assays. (F) Quantification of cGAS-DNA puncta by imaging of MEF cells expressing GFP-tagged human FL-cGAS or  $\Delta$ N160-cGAS after transfection with Cy5-ISD. Representative images are shown in fig. S6. Values shown are means  $\pm$  SD.  $n = 5$  images. (G) cGAMP production in the MEF cells expressing human FL-cGAS or  $\Delta$ N160-cGAS after transfection with ISD or HT-DNA. Values are means  $\pm$  SD.  $n = 3$ . (F) and (G): \*\*\*\* $P < 0.0001$  (multiple  $t$  tests). cGAS expression levels are shown in fig. S5F. Data are representative of at least three independent experiments.

buffer (Fig. 3C) or at 300 mM NaCl (fig. S5C). The enzymatic activity of FL-cGAS in the presence of HT-DNA was stronger than that of  $\Delta$ N146-cGAS in both low-salt buffer (Fig. 3D) and physiological buffer (Fig. 3E).

To investigate the role of the cGAS N terminus in cells, we reconstituted cGAS-deficient MEF cells with human FL-cGAS or  $\Delta$ N160-cGAS (fig. S5F). After transfection with Cy5-ISD, FL-cGAS formed puncta with Cy5-ISD in the MEF cells, whereas  $\Delta$ N160-cGAS formed fewer puncta (Fig. 3F and fig. S6). cGAMP production was also higher in cells stably expressing FL-cGAS than in those expressing  $\Delta$ N160-cGAS upon transfection with 45-bp ISD or HT-DNA (Fig. 3G; cGAS expression levels are shown in fig. S5F).

cGAS enzyme activity was much weaker in the assay with physiological buffer than in the assay with low-salt buffer (fig. S5, D and E). This raises the question of how cGAS is activated in cells. We found that zinc ions substantially promoted the activity of recombinant cGAS in the physiological buffer and that this enhancement could be partially replaced by other ions, such as  $Mn^{2+}$  or  $Co^{2+}$  (Fig. 4A and fig. S7A), which is a general characteristic of enzymes that require zinc (23). Similarly,  $Zn^{2+}$  was the most efficient in activating mouse cGAS (fig. S7, B and C). The optimal con-

centrations of  $Zn^{2+}$  were 160 to 625  $\mu$ M (Fig. 4A and fig. S7C), which is within the physiological concentration range of zinc ions in cells (24, 25).  $Zn^{2+}$  at  $\sim 200$   $\mu$ M markedly facilitated DNA-induced cGAS phase separation at low concentrations of cGAS and DNA (Fig. 4B and fig. S7D). Moreover, cGAS activity at low concentrations was markedly enhanced in the presence of  $Zn^{2+}$  (Fig. 4C). At a concentration as low as 2.5 nM, which is below the concentration of cGAS in the cytoplasm of HeLa cells (approximately 8 to 12 nM) (methods), cGAS underwent phase transition and catalyzed cGAMP synthesis (Fig. 4, B and C). At 10 nM cGAS, 100-bp DNA activated cGAS with a median effective concentration ( $EC_{50}$ ) of  $\sim 1.4$  nM, whereas 45-bp DNA had much weaker activity. DNA at 25 bp or shorter had no detectable activity in the in vitro assay (Fig. 4D). This DNA length-dependent activation of cGAS was largely mirrored in the cellular assay in which a THP-1 reporter cell line was transfected with DNA (Fig. 4E).

Using a thermal shift assay, we found that DNA binding destabilized cGAS but that  $Zn^{2+}$  stabilized the cGAS-DNA complex (Fig. 4F and fig. S8A). Measurements of free  $Zn^{2+}$  concentrations revealed that  $Zn^{2+}$  bound to cGAS but not to DNA (Fig. 4G). To determine whether zinc

plays a role in cGAS activation within cells, we depleted L929 cells of zinc with the zinc-specific chelator  $N,N,N',N'$ -tetrakis(2-pyridinylmethyl)-1,2-ethanediamine (TPEN). Cellular cGAMP production upon transfection with HT-DNA was decreased in a TPEN concentration-dependent manner (Fig. 4H and fig. S8B). Under these conditions, TPEN did not affect the viability of L929 cells (fig. S8C). Live-cell imaging with a zinc-specific fluorescent probe revealed that cGAS-DNA puncta contained zinc (fig. S8, D and E). These results showed that zinc facilitated cGAS activation in cells by promoting cGAS phase transition in the presence of cytosolic DNA.

Thus, DNA binding to cGAS induces a robust phase transition to liquidlike droplets, which function as microreactors in which the enzyme and reactants are concentrated to greatly enhance the production of cGAMP. This mechanism allows cGAS to detect the presence of DNA in the cytoplasm above a certain threshold to trigger a switchlike response. Such a switchlike response is made possible by the multivalent interactions between the DNA binding domains of cGAS and DNA in a manner that depends on the DNA length. This also provides an explanation for why long DNA activates cGAS more efficiently. The binding between cGAS and DNA involves

#### Fig. 4. Zinc ions promote DNA-induced phase separation and activation of cGAS. (A) $Zn^{2+}$

enhances cGAS activation in vitro. Recombinant human FL-cGAS (15 nM) was incubated with ATP, GTP, and DNA in a physiological buffer containing the indicated concentrations of  $Zn^{2+}$ , and cGAMP production was measured.

(B) Quantification of cGAS-DNA condensates in the presence or absence of zinc. Liquid-phase condensates formed after mixing of Alexa Fluor 488-labeled human FL-cGAS with 45-bp Cy3-labeled ISD at the indicated concentrations of each in physiological buffer with or without  $Zn^{2+}$  (200  $\mu$ M). Images were then captured by confocal microscopy, and representative images are shown in fig. S7D. Values are means  $\pm$  SD.  $n = 5$  images.  $P$  values are from multiple  $t$  tests.

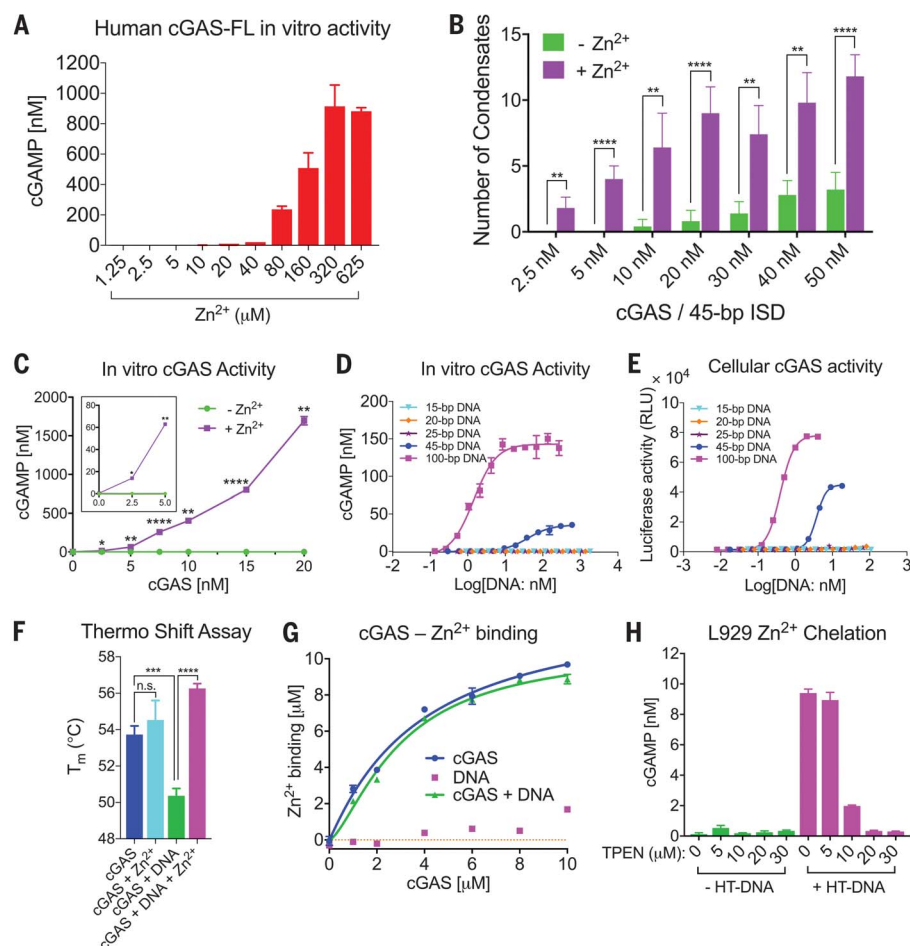
(C) cGAMP production in physiological buffer containing HT-DNA and different concentrations of cGAS in the presence or absence of  $Zn^{2+}$  (200  $\mu$ M). The activity of cGAS at low concentrations is shown in the inset.  $P$  values are from multiple  $t$  tests.

(D) cGAMP production by 10 nM cGAS in physiological buffer containing 200  $\mu$ M  $Zn^{2+}$  and different concentrations of DNA of the indicated lengths.

(E) THP-1-Lucia ISG cells, which harbor a luciferase gene under the ISG54 promoter, were transfected with the indicated DNA for 24 hours, after which the secreted luciferase activity was measured. RLU, relative luciferase units.

(F) Thermal shift assay to measure the stability of cGAS or the cGAS-DNA complex in the presence or absence of  $Zn^{2+}$  (200  $\mu$ M).  $T_m$ , protein melting temperature. Values are means  $\pm$  SD.  $n = 3$ .  $P$  values are from an unpaired  $t$  test.

(G) Measurement of cGAS binding to zinc. Zinc ions (10  $\mu$ M) were incubated with various concentrations of DNA, cGAS, or both, and the solution was passed through a centrifugal filter, after which the zinc ion concentration in the filtrate was measured. The dissociation constant ( $K_d$ ) values for zinc binding to cGAS and the cGAS-DNA complex were  $3.9 \pm 1.3$   $\mu$ M and  $3.0 \pm 0.4$   $\mu$ M, respectively. (H) Depletion of intracellular zinc inhibits cGAS activation by DNA. L929 cells were incubated with the indicated concentrations



of the  $Zn^{2+}$  chelator TPEN for 2 hours before transfection with HT-DNA. cGAMP production was measured by a bioassay. Images depicting intracellular zinc depletion are shown in fig. S8B. (B), (C), and (F): n.s.,  $P > 0.0332$ ; \* $P < 0.0332$ ; \*\* $P < 0.0021$ ; \*\*\* $P < 0.0002$ ; \*\*\*\* $P < 0.0001$ . Error bars represent the variation range of duplicate assays unless otherwise indicated. Data are representative of at least three independent experiments.

extensive ionic interactions between the positively charged surfaces of cGAS and negatively charged DNA. Such interactions are vulnerable to cytoplasmic salt concentrations, which may be a mechanism to prevent spurious activation of cGAS by self-DNA below a certain threshold. However, we found that zinc ions could substantially enhance cGAS phase separation and its enzymatic activation at physiological salt concentrations. Free zinc ions are stored mainly in organelles such as the mitochondria and the endoplasmic reticulum (26), and their delivery to the cytosol may be another avenue by which cGAS activity is regulated in cells. DNA binding to cGAS induced formation of cGAS-DNA condensates, which were observed as cytoplasmic foci within cells. Further characterization of the dynamics and composition of the cGAS condensates should provide deeper insights into the mechanism by which cGAS activity is tightly reg-

ulated to trigger an appropriate immune response to pathogens while simultaneously avoiding autoimmune reactions to self-tissues.

#### REFERENCES AND NOTES

1. L. Sun, J. Wu, F. Du, X. Chen, Z. J. Chen, *Science* **339**, 786–791 (2013).
2. J. Wu et al., *Science* **339**, 826–830 (2013).
3. J. Wu, Z. J. Chen, *Annu. Rev. Immunol.* **32**, 461–488 (2014).
4. G. N. Barber, *Nat. Rev. Immunol.* **15**, 760–770 (2015).
5. Q. Chen, L. Sun, Z. J. Chen, *Nat. Immunol.* **17**, 1142–1149 (2016).
6. L. Andreeva et al., *Nature* **549**, 394–398 (2017).
7. X. Zhang et al., *Cell Rep.* **6**, 421–430 (2014).
8. P. J. Kranzusch, A. S. Lee, J. M. Berger, J. A. Doudna, *Cell Rep.* **3**, 1362–1368 (2013).
9. P. Gao et al., *Cell* **153**, 1094–1107 (2013).
10. F. Civril et al., *Nature* **498**, 332–337 (2013).
11. X. Li et al., *Immunity* **39**, 1019–1031 (2013).
12. C. P. Brangwynne et al., *Science* **324**, 1729–1732 (2009).
13. C. P. Brangwynne, T. J. Mitchison, A. A. Hyman, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 4334–4339 (2011).

14. A. A. Hyman, K. Simons, *Science* **337**, 1047–1049 (2012).
15. A. A. Hyman, C. A. Weber, F. Jülicher, *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).
16. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
17. Y. Shin, C. P. Brangwynne, *Science* **357**, eaaf4382 (2017).
18. T. W. Han et al., *Cell* **149**, 768–779 (2012).
19. M. Kato et al., *Cell* **149**, 753–767 (2012).
20. P. Li et al., *Nature* **483**, 336–340 (2012).
21. Y. Lin, D. S. Protter, M. K. Rosen, R. Parker, *Mol. Cell* **60**, 208–219 (2015).
22. P. Seeman, D. Cheng, G. H. Iles, *J. Cell Biol.* **56**, 519–527 (1973).
23. I. Bertini, H. B. Gray, S. J. Lippard, J. S. Valentine, *Bioinorganic Chemistry* (University Science Books, 1994).
24. D. J. Eide, *Biochim. Biophys. Acta* **1763**, 711–722 (2006).
25. W. Maret, *Metallomics* **7**, 202–211 (2015).
26. Q. Lu, H. Haragopal, K. G. Slepchenko, C. Stork, Y. V. Li, *Int. J. Physiol. Pathophysiol. Pharmacol.* **8**, 35–43 (2016).

#### ACKNOWLEDGMENTS

We thank H. Yang for generating the MEF<sup>cGAS KO</sup>-GFP-cGAS and MEF<sup>cGAS KO</sup>-GFP-ΔN160cGAS cell lines, L. Sun for helping with cGAMP bioassays, C. Zhang for helping with live-cell imaging, and



S. Banani for discussions regarding liquid-liquid phase separation of proteins. **Funding:** This work was supported by grants from the Lupus Research Alliance, the Cancer Prevention and Research Institute of Texas (RP120718 and RP150498), and the Welch Foundation (I-1389). Z.J.C. is an investigator of the Howard Hughes Medical Institute. **Author contributions:** M.D. designed and performed experiments. M.D. and Z.J.C. wrote and revised the manuscript. **Competing interests:** All authors declare no conflicts

of interest. **Data and materials availability:** All data needed to evaluate the conclusions in this paper are present either in the main text or in the supplementary materials.

SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/361/6403/704/suppl/DC1](http://www.sciencemag.org/content/361/6403/704/suppl/DC1)  
Materials and Methods

Figs. S1 to S8  
Table S1  
References (27–30)  
Movie S1

24 January 2018; accepted 27 June 2018  
Published online 5 July 2018  
[10.1126/science.aat1022](https://doi.org/10.1126/science.aat1022)

## TRANSLATIONAL CONTROL

# Fragile X mental retardation 1 gene enhances the translation of large autism-related proteins

Ethan J. Greenblatt and Allan C. Spradling\*

Mutations in the fragile X mental retardation 1 gene (*FMR1*) cause the most common inherited human autism spectrum disorder. *FMR1* influences messenger RNA (mRNA) translation, but identifying functional targets has been difficult. We analyzed quiescent *Drosophila* oocytes, which, like neural synapses, depend heavily on translating stored mRNA. Ribosome profiling revealed that *FMR1* enhances rather than represses the translation of mRNAs that overlap previously identified *FMR1* targets, and acts preferentially on large proteins. Human homologs of at least 20 targets are associated with dominant intellectual disability, and 30 others with recessive neurodevelopmental dysfunction. Stored oocytes lacking *FMR1* usually generate embryos with severe neural defects, unlike stored wild-type oocytes, which suggests that translation of multiple large proteins by stored mRNAs is defective in fragile X syndrome and possibly other autism spectrum disorders.

**F**MR1 is a polysome-associated RNA binding protein required for the nervous system and ovary to develop and function normally in humans, mice, and *Drosophila* (1, 2). Both tissues translationally control stored mRNAs associated with *FMR1*-containing ribonucleoprotein particles (RNPs) (3–6), which suggests that *FMR1* has a specific function in using stored mRNAs. However, the challenge of obtaining highly enriched *FMR1*-containing RNPs from neural tissue may have contributed to difficulties in defining *FMR1* target genes (7–9). We reasoned that *FMR1* function could be studied in a physiologically relevant context using mature *Drosophila* oocytes, which lack transcription and depend entirely on ongoing translation.

*Drosophila* oogenesis is highly amenable to such studies because each female can hold up to 30 mature oocytes per ovary for several weeks (Fig. 1A). In the ovary, each oocyte is surrounded by 800 somatic cells constituting a follicle. We found a reliable method (see supplementary materials) to maintain completed follicles in the ovary for a known period of time in the absence of new follicle maturation (Fig. 1B). Ovulation and fertilization could then be induced by adding males to test the stored oocyte's ability to support embryogenesis.

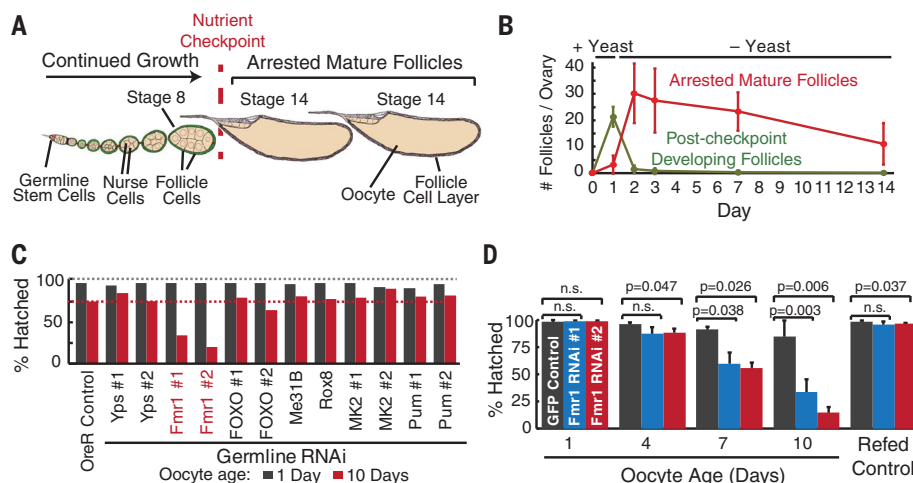
We tested the function of specific genes during oocyte storage by depleting their transcripts with germline-specific GAL4-driven RNA interference (RNAi), which is produced throughout oogenesis starting in the germline stem cell. Disrupting a gene required during oocyte storage would cause oocytes to at first develop normally, and then to lose developmental capacity more rapidly than the wild type during further storage. To screen

for such genes, we depleted *Fmr1* mRNA or several other candidate gene transcripts and analyzed oocytes after a storage period of 1 day or 10 days (Fig. 1C). In most gene knockdown lines and in wild-type controls, the hatch rate was nearly 100% after 1 day of storage, and after 10 days only dropped to 80%. In contrast, *Fmr1*-depleted oocytes hatched normally at first, but after 10 days of storage, only 20 to 30% of them completed development (Fig. 1C). Germline *Fmr1* RNAi dras-

tically reduced *Fmr1* mRNA levels, and antibody staining confirmed that *FMR1* protein was effectively depleted throughout oogenesis specifically in germ cells but not in somatic cells (fig. S1).

We validated the *Fmr1* requirement in stored oocytes by showing that oocyte viability drops continuously over time without *Fmr1* (Fig. 1D). The defects were specific to the old oocytes themselves. Refeeding the mother caused the remaining stored oocytes to be laid and new oocytes to mature, which were fully functional (Fig. 1D). Thus, germline *Fmr1* RNAi does not observably impair stem cells, follicle development, or any nonautonomous aspects of female germline function. This differs from *Fmr1* mutants, which lose stem cells and produce fewer follicles because mutants affect niche cells and other somatic cells in addition to germ cells (10).

We analyzed embryos derived from control and *FMR1*-deficient oocytes to investigate what processes are affected by *FMR1* depletion. Embryos from control oocytes developed a normal nervous system regardless of prior storage, as shown by staining with broadly expressed neural markers (Fig. 2, A and B). The same was true of embryos derived from *Fmr1* RNAi oocytes after 1 day of storage (Fig. 2A). However, in embryos derived from follicles lacking *FMR1* during 10 days of storage, ventral nerve cord-specific labeling showed missing commissures and breaks in the longitudinal connectives (Fig. 2C), in contrast to the wild type. These neuronal defects were not due to a generalized deterioration of the *FMR1*-depleted oocyte's ability to support embryonic development (fig. S2). We never observed comparable neural defects in embryos derived from



**Fig. 1. *Fmr1* is specifically required during the storage of mature, quiescent stage 14 oocytes in the ovary.** (A) Schematic of a *Drosophila* ovariole with immature precheckpoint follicles and two stored mature stage 14 follicles. (B) Arrested mature follicle stability (red) following feeding protocol as described (see supplementary materials). (C) *Fmr1* knockdown (lines 1 and 2), but not knockdowns of controls or other indicated genes, specifically reduces 10-day stored (red) but not 1-day stored (black) oocytes from developing into hatching larvae. (D) *Fmr1* germline RNAi during storage progressively reduces the fraction of mature oocytes competent after 1, 4, 7, or 10 days of storage to support development. Refeeding females to promote maturation of fresh stage 8 follicles restores full developmental potential. Error bars in (B) and (D) denote SD; n.s., not significant. GFP, green fluorescent protein.

Howard Hughes Medical Institute Research Laboratories, Department of Embryology, Carnegie Institution for Science, Baltimore, MD 21218, USA.

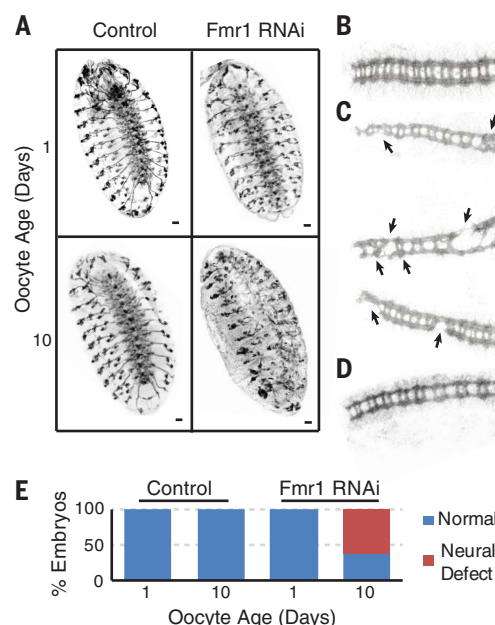
\*Corresponding author. Email: spradling@ciwemb.edu

wild-type oocytes, even when they were stored for 14 days when fewer than 20% of their embryos developed to hatching (Fig. 2D). In contrast, more than 50% of embryos derived from FMR1-deficient oocytes stored in the ovary for 10 days developed a severely abnormal nervous system (Fig. 2E). Thus, disrupting FMR1 function while oocytes are fully dependent on translational regulation specifically compromised their ability to support neural development relative to controls.

To determine how FMR1 disruption affects oocyte translation and to identify FMR1 translational target genes that may be important for sustaining nervous system development, we developed a ribosome-profiling protocol based on (11) (see supplementary materials) to quantify oocyte translation in an unbiased, genome-wide manner. Flies induced for oocyte storage were analyzed after only 1 or 2 days to identify initial translational changes in healthy *Fmr1* knockdown oocytes prior to viability reduction. Although whole ovaries were needed to obtain enough material for profiling, most of the analyzed ribosomes should still derive from stored stage 14 follicles, because they are larger than the total size of earlier follicles. Interestingly, the ribosome footprints and transcript levels of most mRNAs were unaffected by germline *Fmr1* knockdown (Fig. 3A); this result indicates that FMR1 does not generally control translation or mRNA stability.

From 11 independent, highly reproducible ribosome profiling experiments (fig. S3), we identified 421 genes whose germline expression significantly declined and 14 genes that significantly increased expression in *Fmr1* RNAi (Fig. 3, B and C, and tables S1 and S2). Except for *Fmr1* itself, translation of the significantly altered targets generally decreased by a factor of 1.3 to 2.5, which we verified by Western blotting (fig. S4), whereas their RNA levels were unchanged (Fig. 3B). Many down-regulated genes, at least 56 of 421, are orthologs of human genes that have been implicated in human neurodevelopmental syndromes (Fig. 3C and table S3)—a proportion greater than expected by chance ( $P = 1.1 \times 10^{-9}$ ; fig. S5A). For example, the neurofibromatosis gene *Nf1*, which is associated with cognitive and behavioral disorders and neural tumors (12), was reduced by a factor of 2.5. Several E3 ubiquitin ligases, including CTRIP/TRIP12, POE/UBR4, and HUWE1, whose human homologs are associated with intellectual disability, autism, early-onset dementia, and schizophrenia (13–15), were reduced by about a factor of 2. In total, homologs of at least 20 dominant autism/intellectual disability genes were significantly reduced (table S3). Because mutations in these genes are dominant (16), a factor of 2 reduction in expression has potential consequences, even for a single target.

To determine whether FMR1 acted on target transcripts through direct binding, we compared our candidate FMR1 targets with a previous report, which used proximity-based strategies to cross-link mRNAs in brain tissue before immunoprecipitation of FMR1 (7). We found significant overlap between both datasets ( $P = 1.2 \times 10^{-25}$ ; fig. S5B), which suggests that FMR1 directly



**Fig. 2. Stored FMR1-depleted oocytes frequently generate embryos with neural defects.**

(A) Control oocytes stored in vivo for 1 day or 10 days support normal embryonic nervous system development (22C10 antibody). *Fmr1* RNAi oocytes stored in vivo for 1 day support normal development but after storage for 10 days produce highly abnormal nervous system development. (B) Normal ventral nerve cord (BP102 antibody) from embryo developed from 10-day control oocyte. (C) Abnormal ventral nerve cords including broken or fused connectives (arrows) from three embryos developed from 10-day *Fmr1* RNAi oocytes. (D) Normal nerve cord from a control oocyte stored for 14 days. (E) Summary of nervous system development in embryos from control (GFP) and *Fmr1* RNAi oocytes ( $N \geq 50$ ). All scale bars, 20  $\mu$ m.

binds many affected transcripts. However, our results differed from prior studies in two important ways. First, the great majority of targets decreased in expression, indicating that FMR1 usually enhances rather than represses translation, in contrast to most (17, 18) but not all (9, 19) previous reports. This difference might arise because multiple FMR1 targets act negatively on protein stability, translation, or cell growth, including nine ubiquitin ligases, Nf1, and Not1 (table S1). Down-regulation of these negative regulators after *Fmr1* loss might substantially increase protein levels, simulating the direct effects of a translational repressor.

Second, almost all of the affected proteins are much larger in size than the average *Drosophila* protein. Dividing mRNAs undergoing translation into size classes showed that FMR1 strikingly affects translation in proportion to protein size (Fig. 3, C and D) and to some extent UTR length (fig. S6A), but not by transcript level (fig. S6D). The translation of nearly half (46%) of expressed proteins longer than 2000 amino acids, 13% of proteins 750 to 1000 amino acids in length, but only 1% of proteins shorter than 250 amino acids was significantly reduced with *Fmr1* knockdown (Fig. 3C). *Fmr1* RNAi did not always impair the translation of large proteins, as there was a broad response to *Fmr1* knockdown (Fig. 3C) and the translation of many large proteins was only minimally affected. FMR1 targets had low translation efficiencies (TEs) in oocytes (Fig. 3D), similar to long transcripts generally (fig. S6B). However, FMR1 boosted the translation of affected long transcripts regardless of TE (Fig. 3E), implicating size and not low TE as the predominant factor. The size effect was not due to reduced processivity, because we observed a uniform reduction in footprints across the entire coding sequence of target mRNAs (Fig. 3F). The

preferential effect on large mRNAs is likely mediated by direct FMR1 binding, because the average size of target proteins common to both this study and (7) was 1841 amino acids. Furthermore, genes linked to autism as a group are significantly longer than average (16, 20).

The *Poe/Ubr4* gene, encoding one of the longest *Drosophila* proteins (5322 amino acids), was investigated as an FMR1 target with potentially large effects, something not previously identified. Both *Poe* and *Fmr1* mutant *Drosophila* are male sterile, cannot fly, and show increased neuromuscular function synaptic excitability (21–23). Stored *Poe* mutant oocytes lost developmental competence at the same rate as *Fmr1* germline RNAi oocytes (Fig. 4A), and these embryos also developed a high frequency of neural defects (Fig. 4B and fig. S7A). Beginning in nearly mature follicles, POE protein formed 0.5- to 2- $\mu$ m spherical particles in germ cells (Fig. 4C), which were distinct from RNP granules containing the P body marker TRAL (fig. S8). POE antibody staining was lost in *Poe* germline RNAi and *Poe* mutants (Fig. 4D). As predicted, POE protein levels were reduced and particles were reduced or eliminated in *Fmr1* germline RNAi and in *Fmr1*-null egg chambers (Fig. 4D). These observations demonstrate that *Poe* is a major functional target of FMR1, and that POE is itself essential to maintain the oocyte's ability to support neural development. However, overexpression of POE using a duplication in an *Fmr1* RNAi background restored POE granule expression (Fig. 4D) but did not rescue *Fmr1* RNAi lethality upon storage (fig. S7F).

FMR1's function in maintaining an oocyte's ability to support neural development may only be revealed during oocyte storage because some FMR1 targets act catalytically. For example, E3 ligases whose levels are reduced in FMR1-deficient



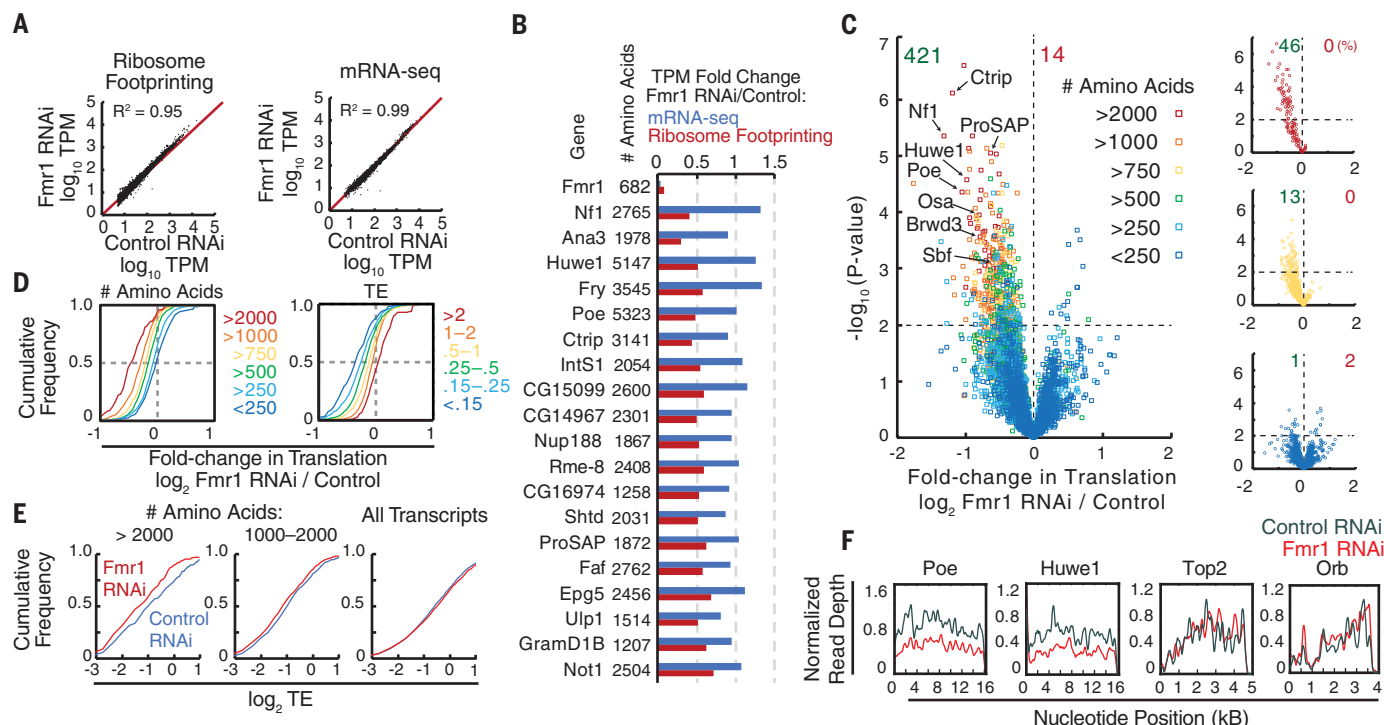
oocytes might no longer be able to prevent the overaccumulation of some of their target proteins, which might eventually reach levels in the stored oocytes that interfere with neural development.

FMR1 regulates translation both in the ovary and during neural development as part of RNP granules that may either stimulate or repress translation. Studies of these granules suggest a

potential explanation for the size effect we observed on target proteins. Mammalian and yeast mRNAs that are large and/or inefficiently translated preferentially associate with stress granules where their translation is repressed (24). FMR1 may function on specific targets to counteract the inherent general tendency of large mRNAs to be segregated into inactive RNP particles. Alterna-

tively, FMR1 might directly or indirectly promote translation initiation in association with RNPs, or it might affect mRNA transport along microtubules to sites of active initiation.

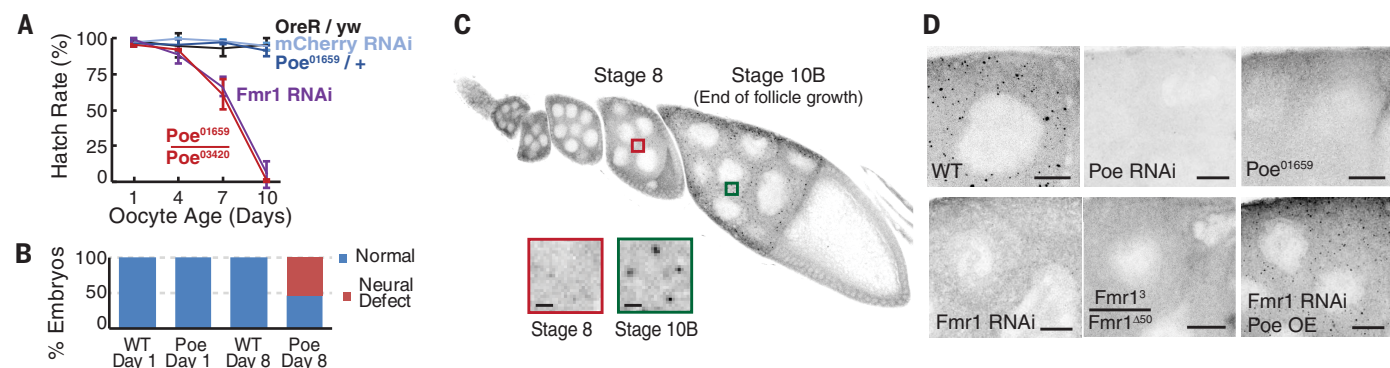
We propose that maintaining the translation of large mRNAs in the translational environments experienced far from the nucleus at synapses in many neurons represents a general challenge that



**Fig. 3. FMR1 stimulates translation during storage of transcripts from multiple intellectual disability and autism genes.** (A) Translational profile [ $\log_{10}(\text{TPM})$ , transcripts per million] and mRNA abundance profile [mRNA-seq,  $\log_{10}(\text{TPM})$ ] are highly similar between control and *Fmr1* RNAi oocytes (stored 1 to 2 days). (B) Top genes translationally reduced by *Fmr1* RNAi from 11 ribosome-footprinting experiments do not show significant changes in mRNA levels. (C) Significance versus relative (fold) change plot reveals 421 candidate targets translationally stimulated by

*FMR1* ( $P < 0.01$ ,  $t$  test). Protein size class is indicated by color.

(D) Cumulative plot of translation (*Fmr1* RNAi/control) as a function of protein size (left) or translational efficiency (TE, right) defined as ribosome-footprinting TPM(*Fmr1* RNAi)/mRNA-seq TPM. (E) Translation of large mRNAs in *Fmr1* RNAi versus controls is reduced independent of TE. (F) Normalized read depth is plotted for two FMR1 targets (*Poe* and *Huwe1*) and two nontargets (*Orb* and *Top2*). In *Fmr1* RNAi oocytes, target gene footprints are reduced at all positions along the mRNA.



**Fig. 4. *Poe* is required for oocyte storage and neural development.**

(A) *Poe* mutation accelerates oocyte decline during storage. Error bars denote SD. (B) *Poe* mutant oocytes frequently fail to support normal neural development after prolonged storage. (C) POE antibody staining (see supplementary materials) during follicle development, showing

germline granules that arise in maturing follicles. Scale bars, 3  $\mu\text{m}$ .

(D) Many POE granules are seen in wild-type stage 10 follicles but not in *Poe* RNAi, *Poe*<sup>01659</sup>, *Fmr1* RNAi, or null *Fmr1*<sup>3/Δ50</sup> follicles. *Fmr1* RNAi combined with POE overexpression (OE) recovers POE granules. Scale bars, 20  $\mu\text{m}$ .

not only underlies fragile X syndrome but is relevant to other neurodevelopmental conditions. The same challenges likely exist in oocytes, spermatocytes, and non-neural somatic cells that require regulated translation from stored mRNAs. Because the challenges of translating large proteins are likely to increase in adult neurons under the influence of aging, the pathways and targets assayed here may contribute to adult-onset neural impairments such as schizophrenia and dementia. Improved knowledge of how FMR1 preserves target translation, and of the identities of major target genes such as *Poe/Ubr4*, may open new opportunities to monitor susceptible cells and to intervene to mitigate declining levels of the most critical targets. Small-molecule agents that counteract the tendency of large mRNAs to be segregated into inactive granules represent potentially valuable therapeutics. Continued study of these highly conserved pathways in *Drosophila* represents one powerful and efficient means to further address both the fundamental and applied implications of these findings.

## REFERENCES AND NOTES

1. R. J. Hagerman *et al.*, *Nat. Rev. Dis. Primers* **3**, 17065 (2017).
2. M. Drozd, B. Bardoni, M. Capovilla, *Front. Mol. Neurosci.* **11**, 124 (2018).
3. S. A. Barbee *et al.*, *Neuron* **52**, 997–1009 (2006).
4. R. Rosario *et al.*, *PLOS ONE* **11**, e0163987 (2016).
5. S. B. Christie, M. R. Akins, J. E. Schwob, J. R. Fallon, *J. Neurosci.* **29**, 1514–1524 (2009).
6. A. Costa *et al.*, *Dev. Cell* **8**, 331–342 (2005).
7. J. C. Darnell *et al.*, *Cell* **146**, 247–261 (2011).
8. M. Ascano Jr. *et al.*, *Nature* **492**, 382–386 (2012).
9. R. Tabet *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, E3619–E3628 (2016).
10. L. Yang *et al.*, *Hum. Mol. Genet.* **16**, 1814–1820 (2007).
11. J. G. Dunn, C. K. Foo, N. G. Belletier, E. R. Gavis, J. S. Weissman, *eLife* **2**, e01179 (2013).
12. S. M. Morris *et al.*, *JAMA Psychiatry* **73**, 1276–1284 (2016).
13. J. Zhang *et al.*, *Hum. Genet.* **136**, 377–386 (2017).
14. D. Monies *et al.*, *Hum. Genet.* **136**, 921–939 (2017).
15. M. Bosshard *et al.*, *Sci. Rep.* **7**, 15050 (2017).
16. W. Perea *et al.*, *Nucleic Acids Res.* **46**, D1049–D1054 (2018).
17. F. Zalfa *et al.*, *Cell* **112**, 317–327 (2003).
18. B. Lagerbauer, D. Ostareck, E. M. Keidel, A. Ostareck-Lederer, U. Fischer, *Hum. Mol. Genet.* **10**, 329–338 (2001).
19. E. G. Bechara *et al.*, *PLOS Biol.* **7**, e1000016 (2009).
20. I. F. King *et al.*, *Nature* **501**, 58–62 (2013).
21. S. Richards, T. Hillman, M. Stern, *Genetics* **142**, 1215–1223 (1996).
22. J. J. Fabrizio, G. Hime, S. K. Lemmon, C. Bazinet, *Development* **125**, 1833–1843 (1998).
23. Y. Q. Zhang *et al.*, *Cell* **107**, 591–603 (2001).
24. A. Khong *et al.*, *Mol. Cell* **68**, 808–820.e5 (2017).

## ACKNOWLEDGMENTS

We thank A. Pinder for help generating RNA-seq libraries, J. Dunn (UCSF) for technical advice for our ribosome footprinting experiments, and E. Wagner (UTMB) and T. Jongs (UPenn) for generously providing the INTS1 antibody and the *Fmr1*<sup>13</sup> mutant fly strain, respectively. **Funding:** Supported by the Jane Coffin Childs Memorial Fund (E.G.) and the Howard Hughes Medical Institute (A.C.S.). **Author contributions:** Both E.G. and A.C.S. contributed to all aspects of this research except that E.G. performed the research. **Competing interests:** The authors declare no competing interests. **Data availability:** Ribosome profiling and mRNA-seq data reported here are available from the NCBI BioProject website under accession number PRJNA466150.

## SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6403/709/suppl/DC1  
Materials and Methods  
Figs. S1 to S8  
Tables S1 to S5  
References (25–27)

13 January 2018; resubmitted 5 May 2018  
Accepted 27 June 2018  
10.1126/science.aas9963

ARTIFICIAL INTELLIGENCE ■ AUTONOMOUS VEHICLES ■ BIOMATERIALS IN ROBOTICS ■ HUMANOIDS ■ LAND & UNDERSEA ROBOTS ■ MEDICAL & SURGICAL ROBOTS  
MICRO/NANO ROBOTS ■ ROBOT ENGINEERING ■ ROBOTS IN EDUCATION ■ SPACE ROBOTS ■ THEORETICAL ADVANCES WITH POSSIBLE APPLICATIONS

# Transforming the **Future** of **Robotics**



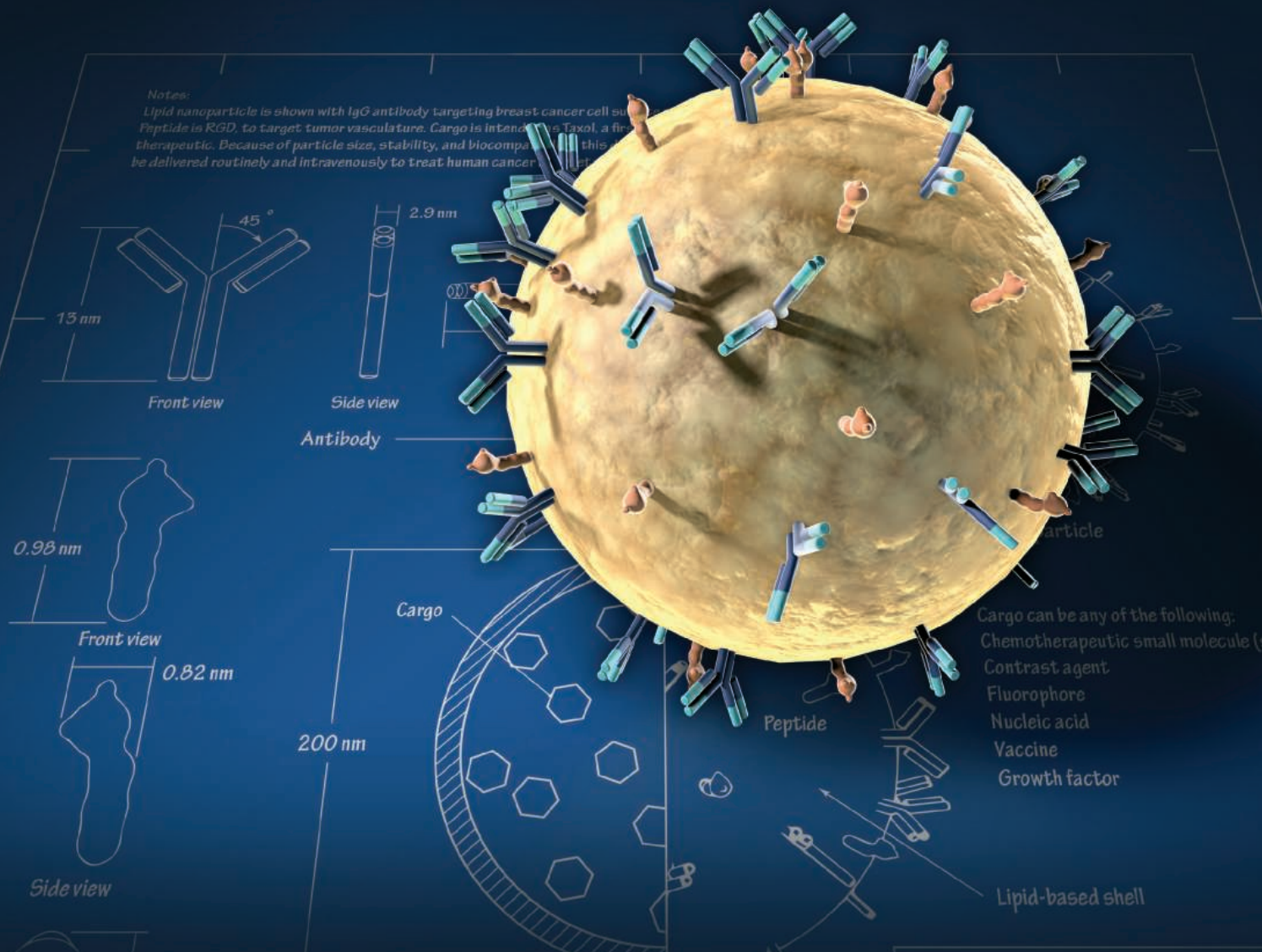
As a multidisciplinary online-only journal, *Science Robotics* publishes original, peer-reviewed, research articles that advance the field of robotics. The journal provides a central forum for communication of new ideas, general principles, and original developments in research and applications of robotics for all environments.

Learn more at: [ScienceRobotics.org](https://www.sciencerobotics.org)

**Science  
Robotics**  
AAAS



# Make Your Research Hit the Target



Science Translational Medicine publishes peer-reviewed, cutting-edge biomedical research in the fields of cardiology, cancer, immunotherapy, infectious diseases and vaccines, bioengineering and devices, neurology and neurodegenerative diseases, obesity, diabetes and metabolic disorders, drug discovery, genomic medicine, imaging, stem cell therapy and regenerative medicine.

**Submit your research today**

Learn more at: [ScienceTranslationalMedicine.org](http://ScienceTranslationalMedicine.org)

Designer Nanomedicine  
**Science  
Translational  
Medicine**  
AAAS



### Arbitrary Waveform Generator

The HDAWG from Zurich Instruments is an arbitrary waveform generator (AWG) with the highest channel density (HD) and shortest trigger latency available.

HDAWG offers up to eight channels with 16-bit output and a sample rate of 2.4 gigasamples per second. A waveform memory of 500 megasamples, 32-bit digital input/output (I/O), and transistor-transistor logic (TTL) marker output complete the package. For applications requiring many channels, multiple instruments can be synchronized and centrally controlled. HDAWG includes the LabOne user interface and multiple application programming interfaces. Sequences can be easily written, edited, and compiled using the embedded scripting language. These sequences can be swiftly transferred over 1 Gigabit Ethernet or USB 3.0, saving time and allowing the user to maintain an overview of complex signal patterns. HDAWG meets the highest R&D requirements, for example, quantum computing applications to produce pulsed sequences with minimal noise, nuclear magnetic resonance, electronic component testing, spectroscopy, and radar/lidar.

#### Zurich Instruments

For info: +41-(0)-44-515-0410  
www.zhinst.com

### Tip Eject Reload Solution

Reload your racks without coming in contact with the tips. Tip Eject provides a clean reload solution with easy one-handed loading. Tip Eject's rigid plastic covering shields tips from the laboratory environment from the top down, keeping contamination out. It comes with clear casing and colored snap cards for easy tip identification. TipEject takes up one-fifth the bench space of a pack of racked tips. The package is recyclable; by reusing the base, plastic use is reduced by up to 44%. When you've got the need for speed and only one free hand, Tip Eject is the solution.

#### Biotix

For info: 888-330-2396  
biotix.com

### Gene Expression Panels

nCounter Gene Expression Panels are developed in collaboration with leading researchers to deliver the most valuable, relevant panel content across applications, providing a highly multiplexed, simple workflow—no library prep or dilutions required, and just 15 min total hands-on time. Address key challenges in immuno-oncology by measuring multianalyte expression levels simultaneously within the tumor, the immune system, and the tumor microenvironment all on one platform, with NanoString's 3D Biology Technology. nCounter Immunology, Inflammation, and Myeloid Panels are all designed to evaluate biomarkers associated with autoimmune diseases and can be customized with up to 30 additional user-defined genes. Furthermore, nCounter Immunology Panels are ideal for assessing infectious disease immune response and for pathogen identification.

#### NanoString Technologies

For info: 888-358-6266  
www.nanostring.com

### Solid-Phase Microextraction Tips

SPE-it Tips are solid-phase microextraction devices configured to permit isolation of analytes from aqueous solutions followed by direct analysis without further sample processing. The fiber tips are solid-phase microextraction fibers embedded in a micropipette tip for easy handling. The fiber coating is composed of either C18 or PDMS/DVB bonded silica particles embedded in a biocompatible polymer. These tips can be manipulated manually or by robotic equipment, and are shipped in a 96-well format tray. The DART SPE-it kit includes a box of 96 SPE-it fibers (C18 or PDMS/DVB), a deep-well plate shaker, two SPE-it holders (each holds 12 fibers), a module for use with dynamic automated range technology (DART) automation hardware, an SPE-it Kit Place Mat with instructions, and 5 × 96 well plates provided with the initial kit. Additional fibers can be purchased separately in eight-pack units.

#### IonSense

For info: 781-484-1043  
ionsense.com/products/consumables/spe\_it\_tips

### Field Fluorometer

The V-Lux Fluorometer is configured to provide high-quality in situ detection of algae, aromatic hydrocarbons, or tryptophan-like fluorescence. It includes three fluorescence channels as well as absorption, turbidity, and temperature channels, which allow corrections from these potential interferences. Applications include monitoring of road and airport apron runoff, bathing waters, and shellfish waters, plus discharge monitoring within the oil and gas sector. The fluorometer is packaged within a small, 50-mm diameter housing of 158-mm length, is rated to 6,000 m, and has integrated antibiofouling protection. It comes with an internal logger and provides real-time data with a choice of data output protocols including Modbus, SDI-12, and other digital formats, and includes quality control channels. It can be used as a handheld instrument, as part of a flow-through system, or deployed from gliders or underwater vehicles.

#### Chelsea Technologies Group

For info: +44-(0)-20-8481-9000  
www.chelsea.co.uk

### Surface Analysis System

Materials researchers, product developers, and failure-analysis scientists seeking to conduct cost-effective, research-level X-ray photoelectron spectroscopy (XPS) studies can now benefit from the Thermo Scientific Nexsa surface analysis system. The system is designed to easily integrate multiple analytical techniques in a single, compact, fully automated surface analysis instrument. Nexsa combines the high throughput and high sensitivity of the K-Alpha XPS system with the multitechnique capabilities of the ESCALAB Xi<sup>+</sup> XPS microprobe. Nexsa users can add complementary techniques, such as Raman spectroscopy, ion scattering spectroscopy, reflected electron energy loss spectroscopy, and UV photoelectron spectroscopy, to generate multiple measurements from the same point on the sample. Integration of multiple techniques allows users to conduct true correlative analysis, unlocking the potential for further advances in microelectronics, ultrathin films, and nanomaterials development.

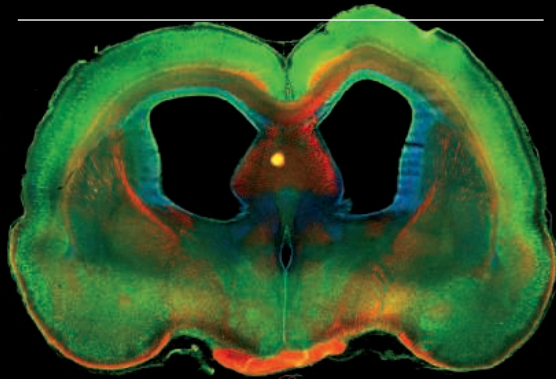
#### Thermo Fisher Scientific

For info: 800-532-4752  
www.thermofisher.com

Electronically submit your new product description or product literature information! Go to [www.sciencemag.org/about/new-products-section](http://www.sciencemag.org/about/new-products-section) for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

# DOES YOUR LAB ANALYZE THE MECHANISMS THAT MEDIATE COMMUNICATION BETWEEN CELLS?



Kong-Yan Wu *et al.* (Zhen-Ge Luo), "Semaphorin 3A activates the guanosine triphosphatase Rab5 to promote growth cone collapse and organize callosal axon projections", *Sci. Signal.* 7, ra81 (2014). Photo Credit: Rat Brain Slice. Photo Credit: Kong-Yan Wu and Zhen-Ge Luo, Chinese Academy of Sciences.

**ScienceSignaling** | AAAS  
CELL SIGNALING IN PHYSIOLOGY AND DISEASE

Find out more about the scope of the journal and submit your research today. [ScienceSignaling.org](http://ScienceSignaling.org)

## Laser-Based Micropipette Puller

- Pull quartz, borosilicate, and aluminosilicate glass
- Fully programmable, including heating filament characteristics
- Pulls electrodes with tips less than 0.03  $\mu\text{m}$
- Optimized velocity sensing circuit
- Long-life  $\text{CO}_2$  laser
- Write-protected for inadvertent changes
- Date/time stamp and total heat time displayed
- P-2000/F is ideal for nanospray and NSOM



## P-2000

The P-2000 integrates a  $\text{CO}_2$  laser-based heat source with the technology derived from our extensive experience with conventional pullers. This system offers capabilities unmatched by other pullers. A significant advance in the technology of fabrication of micropipettes, optical fiber probes, and nanospray tips, is offered with the P-2000 micropipette puller.

**SUTTER INSTRUMENT**

PHONE: +1.415.883.0128 | FAX: +1.415.883.0572  
EMAIL: [INFO@SUTTER.COM](mailto:INFO@SUTTER.COM) | [WWW.SUTTER.COM](http://WWW.SUTTER.COM)

FORCEFORSCIENCE.ORG

# STAND TOGETHER Be a Force for Science

### GET THE FACTS

Understand the science behind the issues that matter.

### FOLLOW AAAS ADVOCACY

Champion public discussion and evidence-based policy.

### TAKE ACTION

Learn ways you can become an advocate and stand up for science.

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE







# Exceptional scientists wanted

Present your work to the world

Are you a representative of the upcoming generation of thought leaders in your field? Together we look forward to your application for the new Sartorius & Science Prize for Regenerative Medicine & Cell Therapy.

Apply now!

[www.passionforscience.com/prize](http://www.passionforscience.com/prize)



The Sartorius & Science  
Prize for Regenerative  
Medicine & Cell Therapy

---

Awarded by



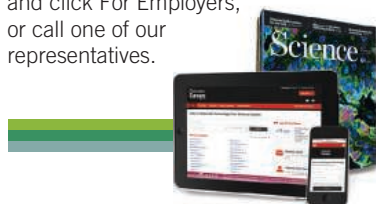
sartorius

Science

# Science Careers

## SCIENCE CAREERS ADVERTISING

For full advertising details, go to [ScienceCareers.org](http://ScienceCareers.org) and click For Employers, or call one of our representatives.



### AMERICAS

+1 202 326-6577  
+1 202 326-6578  
[advertise@sciencecareers.org](mailto:advertise@sciencecareers.org)

### EUROPE, INDIA, AUSTRALIA, NEW ZEALAND, REST OF WORLD

+44 (0) 1223 326527  
[advertise@sciencecareers.org](mailto:advertise@sciencecareers.org)

### CHINA, KOREA, SINGAPORE, TAIWAN, THAILAND

+86 131 4114 0012  
[advertise@sciencecareers.org](mailto:advertise@sciencecareers.org)

### JAPAN

+81 3-6459-4174  
[advertise@sciencecareers.org](mailto:advertise@sciencecareers.org)

## CUSTOMER SERVICE

### AMERICAS

+1 202 326-6577  
**REST OF WORLD**  
+44 (0) 1223 326528

[advertise@sciencecareers.org](mailto:advertise@sciencecareers.org)

All ads submitted for publication must comply with applicable U.S. and non-U.S. laws. *Science* reserves the right to refuse any advertisement at its sole discretion for any reason, including without limitation for offensive language or inappropriate content, and all advertising is subject to publisher approval. *Science* encourages our readers to alert us to any ads that they feel may be discriminatory or offensive.

**ScienceCareers**

FROM THE JOURNAL SCIENCE AAAS

[ScienceCareers.org](http://ScienceCareers.org)



RADCLIFFE INSTITUTE  
FOR ADVANCED STUDY  
HARVARD UNIVERSITY

## Academic Fellowships

The Radcliffe Institute Fellowship Program at Harvard University welcomes fellowship applications in natural sciences and mathematics. The Radcliffe Institute for Advanced Study provides scientists the time and space to pursue their career's best work. At Radcliffe you will have the opportunity to challenge yourself, meet and explore the work of colleagues in other fields, and take advantage of Harvard's many resources, including the extensive library system. The Radcliffe Institute Fellowship Program invites applications from people of all genders, and from all countries. We seek to build a diverse fellowship program. Scientists in any field who have a doctorate in the area of the proposed project (by December 2017) and at least two published articles or monographs are eligible to apply for a Radcliffe Institute fellowship. The stipend amount of \$77,500 is meant to complement sabbatical leave salaries of faculty members. Residence in the Boston area and participation in the Institute community are required during the fellowship year.

**Applications for 2019–2020 are due by October 4, 2018.**

For more information, please visit [www.radcliffe.harvard.edu](http://www.radcliffe.harvard.edu) or email [sciencefellowships@radcliffe.harvard.edu](mailto:sciencefellowships@radcliffe.harvard.edu).

## Search more jobs online

Access hundreds of job postings on [ScienceCareers.org](http://ScienceCareers.org).

**Expand your search today.**



## Advance your career with expert advice from Science Careers.



**Download Free Career Advice Booklets!**  
[ScienceCareers.org/booklets](http://ScienceCareers.org/booklets)

### Featured Topics:

- Networking
- Industry or Academia
- Job Searching
- Non-Bench Careers
- And More



**ScienceCareers**

FROM THE JOURNAL SCIENCE AAAS





# Faculty Positions at the Sloan Kettering Institute Memorial Sloan Kettering Cancer Center

## OPPORTUNITIES IN BIOMEDICAL RESEARCH

- **Cancer Genetics**
- **Cell Signaling & Regulation**
- **Chemistry & Chemical Biology**
- **Computational Biology**
- **Developmental Biology**
- **Experimental Therapeutics**
- **Genome Integrity, Gene Expression, and Epigenetics**
- **Immunology**
- **Stem Cell Biology & Regenerative Medicine**
- **Structural Biology**



## JOIN OUR FACULTY

Successful candidates will hold an appointment in one of SKI's research programs.

Candidates may apply to up to two programs:

### Cancer Biology & Genetics

Chair: Scott Lowe, PhD

### Cell Biology

Chair: Kristian Helin, PhD

### Chemical Biology

Chair: Derek Tan, PhD

### Computational and Systems Biology

Chair: Dana Pe'er, PhD

### Developmental Biology

Chair: Kathryn Anderson, PhD

### Immunology

Chair: Alexander Rudensky, PhD

### Molecular Biology

Chair: John Petrini, PhD

### Molecular Pharmacology

Chair: David Scheinberg, MD, PhD

### Structural Biology

Chair: Nikola Pavletich, PhD



## RESEARCH AND TRAINING

- 130 research laboratories housed in state-of-the-art buildings
- Core facilities offering cutting-edge scientific services
- Over 800 pre- and post-doctoral trainees
- Appointments in the Gerstner Sloan Kettering Graduate School of Biomedical Sciences and the Weill Cornell Graduate School of Medical Sciences

Visit [www.ski.edu](http://www.ski.edu) to learn more.

Application deadline: **October 15, 2018**

Apply at: <https://facultysearch.ski.edu>







**Department of Genetics  
Yale University School of Medicine**



**Professor of Genetics and  
Director of the Center for Genomic Health**

The Department of Genetics at Yale University School of Medicine is searching for a Professor of Genetics with an outstanding record of transformative scientific achievements in Human Genetics and Genomics. We expect that the candidate will lead a vigorous cross-disciplinary research program focused on identifying and characterizing genetic drivers of human disease. As a leader of human genetics both within the department and across the Yale School of Medicine, the successful candidate will have the opportunity to recruit other human geneticists to the Genetics Department and lead a new program in precision medicine as the Scientific Director of the Yale Center for Personalized Medicine and Genomic Health.

We are looking for a dynamic, internationally recognized scientist (Ph.D., M.D., or M.D./Ph.D.) with an outstanding research record of scientific discoveries, as well as a strong track record of training the best innovators in the field of human genetics and genomics.

To apply, please submit your CV to <http://apply.interfolio.com/45539> to the attention of **Antonio Giraldez**, Chair of Genetics. Applications will be reviewed until the position is filled. Inquiries should be addressed to [neltja.brewster@yale.edu](mailto:neltja.brewster@yale.edu).

**Tenure-Track Faculty Position  
(Assistant/Associate/Full Professor)**

The Department of Genetics at Yale University School of Medicine invites applications for junior or senior tenure-track faculty positions. The search is open to investigators from all areas of biological and biomedical research. We are particularly interested in applicants working in one of the following areas: Developmental Biology, Imaging, Quantitative Biology, Computational Biology, Genomics, Systems Biology, and Genetics. Applications from investigators working at the interface of these areas will be strongly considered. The rank of the appointment will be commensurate with experience and the positions come with a substantial start-up package.

The Department of Genetics comprises an exceptional group of 31 primary basic science faculty with research interests including fundamental aspects of Developmental Biology, Genetics, Genomics and Epigenetics, using different model systems including flies, worms, fish and mouse, and humans (<https://medicine.yale.edu/genetics/>). The Department is closely associated with science initiatives at Yale including The Cancer Center, The Center for Neuroscience, The Stem Cell Center and the Yale Center for Genome Analysis.

Candidate must hold a Ph.D., M.D., or equivalent degree. Applicants should upload a cover letter, a curriculum vitae, a description of previous research (1 page), a concise statement of research plans (up to 2 pages), reprints of 2 publications, and the names of 3 references to the Interfolio website at: <https://apply.interfolio.com/52614>. Specific inquiries about the position may be sent to the attention of **Dr. Antonio Giraldez**, Chair of the Department of Genetics, at [genetics.admin@yale.edu](mailto:genetics.admin@yale.edu). Applications will begin to be evaluated on **November 1, 2018**.

Interviews will take place as part of a multidisciplinary symposium including candidates for different searches. Please reserve the dates of January 15, January 22 and February 11 (snow date) as potential dates for the symposium in case you are selected for an interview.

*Yale University is an Affirmative Action/Equal Opportunity Employer. Yale values diversity among its students, staff, and faculty and strongly welcomes applications from women, persons with disabilities, protected veterans, and underrepresented minorities.*



**ICYS Research Fellow at ICYS, NIMS, Japan**

International Center for Young Scientists (ICYS) of National Institute for Materials Science (NIMS) is now seeking a few research fellows who have passion to conduct innovative materials research based on their original idea. ICYS research fellows are expected to pursue high quality research with their initiatives using the most advanced facilities in NIMS.

ICYS offers a "Melting Pot" environment mixing different materials research fields and cultures. Common language in ICYS is English. Clerical and technical support in English will be given by the ICYS staffs. An annual salary of approximately 5.35 million yen will be offered, which may depend on the qualification and performance of the research fellow. Additional research grant of 2 million yen per year will be supplied to each research fellow. The initial contract term is two years and may be extended by another year depending on one's performance.

All applicants must have obtained a PhD degree within the last ten years. Applicants should submit an **application form** including a research proposal to be conducted during the ICYS term, **CV Header**, **CV** with a list of publications and patents (be sure to attach the header), **a list of DOI of journal publications** following our instruction, **PDF files of three significant publications** to ICYS Recruitment Desk by **September 27, 2018 JST**. The format for **application form** and **CV header** can be downloaded from **our website**. Please visit our website for more details.

ICYS Recruitment Desk,  
National Institute for Materials Science  
<http://www.nims.go.jp/icys/recruitment/index.html>



**OPEN FACULTY POSITION  
INSTITUTE OF MOLECULAR BIOLOGY  
ACADEMIA SINICA, TAIWAN, ROC**

One tenure-track faculty position is open for a highly qualified individual to establish independent research programs in **all disciplines of molecular and cellular biology**. Applicants should hold a Ph.D. degree or its equivalent, with appropriate postdoctoral research experience. The successful recruit will be appointed at the levels of **Assistant, Associate, or Full Research Fellows** (equivalent to academic ranks of Assistant, Associate and Full Professors at universities), and receive a generous multi-year start-up package, followed by annual intramural support.

The Institute of Molecular Biology at Academia Sinica (<http://www.imb.sinica.edu.tw/en>) provides an active and stimulating research environment, is well supported by both extramural and long-term intramural funding, and features several core facilities (cell imaging, microarray, Next Generation Sequencing, RNAi, electrophysiology, FACS, bioinformatics and mouse facilities) that provide state-of-the-art resources and key technical expertise to the Institute's research community. Currently two Ph.D. programs, with one recruiting international students, are formally affiliated with the Institute. English is the official language for seminars and lectures at the Institute, and proficiency in Chinese language is not a prerequisite for application.

Applicants should send electronic files of their Curriculum Vitae, a description of past research accomplishments and future research plans, and arrange for three letters of recommendation to be sent directly to **Director Soo-Chen Cheng c/o Ms. Vivi Chiang** ([vivi@imb.sinica.edu.tw](mailto:vivi@imb.sinica.edu.tw)).

The selection process will start on **November 15, 2018** until the positions are filled. Further information can be obtained from **Ms. Vivi Chiang** at [vivi@imb.sinica.edu.tw](mailto:vivi@imb.sinica.edu.tw)

# Director, Center for Immunobiology

Northwestern University Feinberg School of Medicine invites applications and nominations for the inaugural Director, Center for Immunobiology. The new Director will report directly to the Dean of the medical school. In this role, the Director is responsible for oversight of the research and administrative affairs of the Center at the Feinberg School of Medicine.

The Center for Immunobiology is an interdepartmental center with faculty who are engaged in a broad spectrum of cutting-edge basic, translational, and clinical immunology-related research. Principal investigators appointed through the Feinberg School of Medicine are supported by \$471 million of annual research funding. The medical school is embedded among three U.S. News & World Report Honor Roll hospitals; Northwestern Memorial Hospital is ranked 13th, Lurie Children's Hospital 7th, and the Rehabilitation Institute of Chicago 1st.

Successful candidates will possess a MD, PhD, or MD/PhD degree and be eligible for a faculty appointment as a full-time Professor with scholarly accomplishments and interest in recruiting and developing a strong scientific team. Candidate should be recognized investigators in human immunology with complementary strengths in basic and translational investigation.

Email nominations and CVs of appropriate candidates to Ila Allen, recruitment coordinator, at:

[immunobioresearch@northwestern.edu](mailto:immunobioresearch@northwestern.edu)

Apply online at:  
<https://facultyrecruiting.northwestern.edu/apply/MTQz>

Northwestern University is an Equal Opportunity, Affirmative Action Employer of all protected classes, including veterans and individuals with disabilities. Women, racial and ethnic minorities, individuals with disabilities, and veterans are encouraged to apply. Hiring is contingent upon eligibility to work in the United States.



LMU Munich is one of the leading research universities in Europe. With a 500-year history of scholarly excellence, LMU boasts state-of-the-art facilities, innovative research and transdisciplinary co-operation across all fields, from the sciences to the humanities. LMU is based in Munich - and is an important part of Germany's vibrant and dynamic science and technology hub.

To provide outstanding junior researchers in all subject areas with a long-term career perspective, LMU is offering **TENURE TRACK PROFESSORSHIPS TO SUCCESSFUL ERC STARTING GRANTEES** of the European Research Council (for candidates 2-7 years after their PhD).

If you are interested in joining LMU's research community, please contact the appropriate faculty member in your field of research as soon as possible. For further information regarding this offer, please see

[www.lmu.de/excellent/erc-tenuretrack](http://www.lmu.de/excellent/erc-tenuretrack)



By Jeffrey Maloy

# The universe in a classroom

**T**he universe is an unfathomably large place. But from the front of a lecture hall, it can feel suffocatingly tiny. Standing behind a podium at 8 a.m., I looked out into an auditorium at 360 students; 720 eyes staring intently back at me; 3600 fingers furiously pounding away at keyboards, transcribing my uncertainty for posterity. Did I say “polymerase” when I meant to say “primase”? Was my answer to that question clear? The heat radiating from the projector felt stifling. With 5 minutes remaining, I fumbled an explanation. My chest tightened as I scanned the lecture hall, searching for an escape but finding only four walls that seemed to be inching closer. An eternity elapsed in 30 seconds. When I heard a chuckle (or was it a snicker?) from the corner of the room, I yielded to my inadequacy and dismissed class with a precious 2 minutes to spare, promising to rehash the topic at the beginning of the next class.

Later that day, I recounted the gaffe to a colleague. “We’ve all been there,” she said sympathetically. Rationally, I knew she was being earnest. Yet I couldn’t bring myself to truly believe it. What if my mistakes reflect a real deficiency as a teacher, a scientist, a scholar? Unlike her, I don’t have years of experience to validate my ability. The Ph.D. diploma that sits on top of my bookshelf is less than a year old. It was only my second time teaching my own large lecture course to undergraduates.

Ever since my first experience as a teaching assistant in graduate school, I’ve looked forward to the day I would be the person standing behind that podium. Now that I’m there, I can’t shake the feeling that some cosmic accident delivered me. What qualifies me to stand in front of a classroom and explain the mysteries of biology to hundreds of eager students? Three recently earned letters behind my name? A carefully ironed blue Oxford?

In my classes, I preach the gospel of facts and data. But intellectual knowledge and emotional conviction don’t always match up. I’ve spent immeasurable hours convincing my students and peers that they are qualified, that they are good enough, that they are extraordinary. I worry that I may be the exception. Am I a qualified biologist? Am I a good enough mentor? Am I an extraordinary teacher? My degrees and teaching awards should theoretically provide some evidence of my abilities, but emotional conviction remains elusive.

A few days later, as I lay in bed dwelling on my resume of



***“What qualifies me to stand in front of a classroom [of] hundreds of eager students?”***

stage. But I cannot brush off the small dent my existence has made in the experience of one student. One student who sat down at her computer on a Friday evening to say that I made a difference.

The next morning at 8 o’clock, 360 students trickled into the classroom, the groan of folding plastic chairs announcing their arrival. I exhaled and allowed the clamor of the lecture hall to wash over me and drown out my doubts. I plugged my laptop into the projector and walked to the back of the classroom to greet the students. ■

*Jeffrey Maloy is a lecturer and education research fellow at the University of California, Los Angeles. Do you have an interesting career story that you would like to share? Send it to [SciCareerEditor@aaas.org](mailto:SciCareerEditor@aaas.org).*